New Analytic Techniques for Proving the Inherent Ambiguity of Context-Free Languages

Florent Koechlin ✉
Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

Abstract

This article extends the work of Flajolet [10] on the relation between generating series and inherent ambiguity. We first propose an analytic criterion to prove the infinite inherent ambiguity of some context-free languages, and apply it to give a purely combinatorial proof of the infinite ambiguity of Shamir's language. Then we show how Ginsburg and Ullian's criterion on unambiguous bounded languages translates into a useful criterion on generating series, which generalises and simplifies the proof of the recent criterion of Makarov [21]. We then propose a new criterion based on generating series to prove the inherent ambiguity of languages with interlacing patterns, like $\{a^n b^m a^p b^q \,|\, n \neq p \text{ or } m \neq q, \text{ with } n, m, p, q \in \mathbb{N}^*\}$. We illustrate the applicability of these two criteria on many examples.

2012 ACM Subject Classification Theory of computation → Grammars and context-free languages

Keywords and phrases Inherent ambiguity, Infinite ambiguity, Ambiguity, Generating series, Context-free languages, Bounded languages

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2022.41

1 Introduction

A context-free grammar G is said to be *unambiguous* if for any word w recognized by G, there exists exactly one derivation tree for w. A context-free language is called *inherently ambiguous* if it can not be recognized by any unambiguous grammar. Proving that a language is inherently ambiguous is a difficult question, as it is an impossibility notion, and it is undecidable in general [14, 15]. In practice, three different methods have emerged to prove the inherent ambiguity of some context-free languages: an approach based on iterations on derivation trees [25, 26], an other based on iterations on semilinear sets [14, 16, 29], and finally an approach based on generating series [10, 17, 21, 28]. The first two approaches are best suited for (and for the second, limited to) bounded languages.

In this article, we provide new sufficient criteria to prove inherent (infinite) ambiguity, answering two questions of Flajolet [10]. Our main result is an interpretation, in the world of generating series of Ginsburg and Ullian's criteria [14] on semilinear sets. It rediscovers and generalises the criterion recently developped by Makarov [21], while opening the way to new techniques to prove the inherent ambiguity of unbounded languages or bounded languages with an interlacing pattern. In a different direction, we also provide a criterion for inherent infinite ambiguity.

1.1 Motivation and background

Deciding if a grammar is ambiguous is undecidable [6], as well as deciding if a context-free language is inherently ambiguous [14, 15]. However, detecting ambiguity in context-free grammars has strong implications for compilers and parsers. Therefore, identifying inherently ambiguous languages is an important step towards our understanding of the limits of the model of context-free languages to describe natural or programming languages. Let us start with some context on the methods developed so far to establish the inherent ambiguity of a language.

42nd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2022).
Editors: Anuj Dawar and Venkatesan Guruswami; Article No. 41; pp. 41:1–41:22
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Bounded languages. The first techniques developed to prove inherent ambiguity dealt with bounded languages. A language L is called *bounded* if there exist words $w_1, \ldots, w_d$ with $d \geq 1$ such that $L \subseteq w_1^* \ldots w_d^*$. Despite its apparent simplicity, the class of bounded languages is rich enough to provide a large variety of inherently ambiguous languages; furthermore it is often possible to deduce the inherent ambiguity of a context-free language from the inherent ambiguity of a bounded language, using the stability of unambiguous context-free languages under intersection with a regular language [14].

Iteration on derivation trees. In 1961, Parikh was the first to exhibit an inherently ambiguous context-free language, the bounded language $L = \{a^n b^m a^p b^q \,|\, n = p \text{ or } m = q, \text{ with } n, m, p, q \in \mathbb{N}_{>0}\}$ (see [26] and [27, Theorem 3]). Parikh's proof relies on an iteration argument over the derivation trees of any unambiguous grammar recognising L. A few years later, Ogden generalised this method and published his famous lemma [25], which drastically simplified the identification of iterating pairs in derivation trees. Since then, these iterations techniques have been very popular to study several inherently ambiguous languages (see for instance [7, 24, 30, 32]). However, they remain subtle and difficult to set up in general; hence they are sometimes unsuitable to study complex context-free languages.

Iteration on semilinear sets. In 1966, after Parikh's article but before Ogden's lemma, Ginsburg and Ullian succeeded in using strong iterations arguments on derivation trees to characterise exactly the inherent ambiguity of *bounded* context-free languages in terms of their associated semilinear sets [14]. Their result made it possible to prove the inherent ambiguity of bounded languages using iterations on semilinear sets instead of derivation trees [14, 16, 29]. Unfortunately, iterations on semilinear sets turned out to be almost as laborious as on derivation trees. The simplicity of the proof and the strong applications of Ogden's lemma severely contrasted with Ginsburg and Ullian's criterion[1] that was complex to use and required a lot of case analysis. It may explain why iterations on semilinear sets were supplanted by iterations on derivation trees.

Generating series method. In 1987, Flajolet [10] proposed a conceptually new approach, based on generating series and the contraposition of the Chomsky-Schützenberger theorem [6]. The generating series of a language L is the formal series $\sum_n \ell_n x^n$ where ℓ_n denotes the number of words of length n in L. Flajolet's idea consists in showing that a language is inherently ambiguous by computing its generating series – which is a purely combinatorial question, for which there are many techniques [11] – and showing that this series is not algebraic – for which there are also several mathematical characterisations [10]. This method turned out to be very successful, as Flajolet was able to easily prove the inherent ambiguity of a dozen languages in his article. It complemented very well the previous techniques used for proving ambiguity: whereas iterations arguments are rather efficient and fast for proving the inherent ambiguity of languages with a simple structure, which tend to have an algebraic generating series[2], on the opposite side, the generating series approach allows to deal with complex languages that have a transcendental (*i.e.* non-algebraic) generating series and seem out of reach of iterations techniques.

[1] In his book [12, p.211], Ginsburg wondered whether there was a simpler technique to prove the inherent ambiguity of $L := \{a^n b^m c^p \,:\, n = m \text{ or } m = p\}$, and in a sense Ogden answered in the positive.

[2] For example, all bounded context-free languages have a rational generating series

Limits. Nevertheless, the three presented approaches sometimes fail on very simple context-free languages expected to be inherently ambiguous, like the language $L' := \{a^nb^mc^p : n \neq m \text{ or } m \neq p\}$. It has a rational – hence algebraic – generating series, and iterations arguments (whether on trees or semilinear sets) struggle to handle the inequality condition that does not constrain anymore the form of iterating pairs. It is not very surprising that those methods do no cover every language, as hinted by the fact that deciding inherent ambiguity is undecidable.

1.2 Problem statement and contributions

At the end of his article [10], Flajolet raised several open questions about the relation between inherent ambiguity and generating series: is it possible to capture the inherent infinite ambiguity of some context-free languages using analytic tools on generating series? Can rational generating series still be useful to prove the inherent ambiguity of languages like $L' = \{a^nb^mc^p : n \neq m \text{ or } m \neq p\}$? Recently, Makarov [21] answered the second question by using new ideas coming from the generating series of $GF(2)$ grammars. He provided a simple criterion on rational series to prove the inherent ambiguity of some bounded languages on $a_1^* \dots a_d^*$, where the a_i's are distinct letters, and proved the inherent ambiguity of L'.

In this article, we give new answers to the two open questions of Flajolet about inherent ambiguity and infinite inherent ambiguity. We first propose an analytic technique to prove the infinite inherent ambiguity of context-free languages (Theorem 4), and apply it to give a *purely combinatorial* proof of the infinite ambiguity of Shamir's language (Corollary 7). Then we use Ginsburg and Ullian's characterisation to derive a simple criterion (Theorem 12) on generating series to prove the inherent ambiguity of some bounded languages, which both generalises and simplifies the proof of an analogous criterion recently found by [21]. We then propose a new criterion based on generating series to prove the inherent ambiguity of languages with an interlacing pattern, that are not covered by [21], like $L'' = \{a^nb^ma^pb^q \,|\, n \neq p \text{ or } m \neq q, \text{ with } n,m,p,q \in \mathbb{N}^*\}$ (Theorem 21). To make them amenable to the wider audience possible, these criteria only require a basic knowledge in combinatorics and in polynomials in several variables.

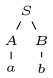
1.3 Related work

To the author's knowledge, since Flajolet's article, and until Makarov's new criterion [21], no real new successful approach based on generating series has been proposed to prove the inherent ambiguity of languages. Several years after Flajolet's article, a subclass of unambiguous context-free language (called slender languages) has been shown to be associated to rational series [17], but their criterion can be in fact interpreted as a shortcut of Flajolet's technique[3]. More recently [1], the class of generating series associated to unambiguous context-free grammars, called $\mathbb{N}$-algebraic series, has been precisely described as well as their asymptotic behaviour. This class of generating series does, however, enjoy less closure properties than algebraic series, which makes them less applicable for proving inherent ambiguity.

If the techniques developed in this article are based on generating series and hence lie in the continuity of Flajolet's method [10], they can also be seen as a nice alliance of the three historical techniques presented in this introduction: we use Flajolet's idea to study ambiguity through generating series [10], in order to revisit from this point of view Ginsburg and Ullian's criteria [14], whose proof relies on iterations in derivation trees.

[3] By [2], if the generating series of a slender language is not rational then it is also not algebraic

2 Preliminaries

Context-free languages. A context-free grammar (CFG for short) is a tuple $G = (N, \Sigma, S, D)$, where Σ is a finite set of terminal symbols, N is a finite set of non-terminal symbols, $S \in N$ is the axiom, and $D \subseteq N \times (N \cup \Sigma)^*$ is the finite set of derivation rules. A rule $(A, w) \in D$ is usually written $A \to w$, with $A \in N$ and $w \in (N \cup \Sigma)^*$. The derivation rules of D can be seen as rewriting rules affecting only non terminal symbols. Let w, w' be two words in $(N \cup \Sigma)^*$. The application of a rewriting rule of D to a non-terminal symbol of w is called a derivation step of G from w. If w' is derived from w after one derivation step, we write $w \to_G w'$. A derivation from w to w' is a (possibly empty if $w = w'$) sequence of consecutive derivation steps $w \to_G w_1 \to_G \ldots \to_G w'$, denoted by $w \to_G^* w'$. As the order of the application of the derivation rules is not canonical, a derivation is rather described as a tree, called a *derivation tree*. For instance, if $D = \{S \to AB, A \to a, B \to b\}$, the derivations $S \to AB \to aB \to ab$ and $S \to AB \to Ab \to ab$ have the same derivation tree  and can be identified. A word is called terminal if it contains only terminal symbols. The language of G, denoted by $\mathcal{L}(G) \subseteq \Sigma^*$, is the set of terminal words that can be derived from the axiom S. The grammar G is said to be *unambiguous* if for any word $w \in \mathcal{L}(G)$, there exists exactly one derivation tree for w. A *context-free language* (CFL) is a language recognized by a CFG. A CFL is called *inherently ambiguous* if it is not recognisable by any unambiguous CFG.

Univariate series. Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ be the set of non-negative integer, $\mathbb{Q}$ the set of rational numbers, $\mathbb{F}_2$ the field with two elements, and K an arbitrary field (in practice, $K = \mathbb{Q}$ or $K = \mathbb{F}_2$ in this article). The set of polynomials with coefficients in K and indeterminate x is denoted by $K[x]$. We denote by $K[[x]]$ the set of formal series with coefficients in K, which is the set of infinite polynomials of the form $\sum_{n \in \mathbb{N}} a_n x^n$, with $a_n \in K$. We recall that $(K[[x]], +, \cdot)$ has a ring structure, with respect to the addition and the Cauchy product. The series $S(x) = \sum_{n \in \mathbb{N}} x^n$ satisfies the equation $(1 - x)S(x) = 1$, and is hence written $\frac{1}{1-x}$. The set $K(x)$ denotes the set of rational fractions, which is formally the set of fractions of the form $p(x)/q(x)$ where p, q are both polynomials in $K[x]$, with $q(x) \neq 0$. A univariate series $f(x) \in K[[x]]$ is *rational* if it satisfies an equation of the form $q(x)f(x) = p(x)$, where $p, q \in K[x]$, $q \neq 0$. In this case $f(x)$ is written $p(x)/q(x)$. It is called *algebraic* over K if there exists a non null polynomial $P(x, Y)$, with coefficients in K, such that $P(x, f(x)) = 0$.

Let $n \in \mathbb{N}$. If L is a language, we define ℓ_n the number of words in L of length n. The generating series $L(x)$ of L is the formal series $L(x) := \sum_{n \in \mathbb{N}} \ell_n x^n \in \mathbb{Q}[[x]]$. If G is a CFG recognising L, we denote by g_n the number of derivation trees of terminal words of length n. If g_n is finite for all $n \in \mathbb{N}$, the generating series of the derivation trees of G, defined by the formal series $G(x) := \sum_{n \in \mathbb{N}} g_n x^n$, is well-defined. In this case, the description of the grammar G translates directly into a polynomial system satisfied by $G(x)$, which implies that $G(x)$ is algebraic over $\mathbb{Q}$. If G is unambiguous, then $G(x)$ is well-defined and coincides with $L(x)$, so the generating series of an unambiguous CFL is algebraic over $\mathbb{Q}$: this is the Chomsky-Schützenberger theorem [6]. The subset of series that are the generating series of the derivation trees of a CFG is called the set of $\mathbb{N}-$algebraic series, and it is strictly included in the set of algebraic series over $\mathbb{Q}$ [1].

Multivariate polynomials. For every $d \in \mathbb{N}_{>0}$, $\mathbb{N}^d$ denotes the set of vectors with d coordinates in $\mathbb{N}$. A vector $(v_1, \ldots, v_d) \in \mathbb{N}^d$ will be freely written in a condensed notation $\boldsymbol{v}$. Similarly, the tuple of d variables $(x_1, \ldots, x_d)$ is written $\boldsymbol{x}$. The notation $K[\boldsymbol{x}]$ denotes the ring of multivariate polynomials with indeterminates $\boldsymbol{x} = (x_1, \ldots, x_d)$ and coefficients in K. For $\boldsymbol{v} = (v_1, \ldots, v_d) \in \mathbb{N}^d$, the monomial $x_1^{v_1} \ldots x_d^{v_d}$ is written $\boldsymbol{x}^{\boldsymbol{v}}$. The total degree of $x_1^{v_1} \ldots x_d^{v_d}$ is the number $v_1 + \ldots + v_d$. A polynomial is called *homogenous* if its monomials have the same total degree (for instance $x^2 + xy$ is homogenous but $1 + xy$ is not). A polynomial is called *irreducible* if it is non constant and cannot be decomposed as the product of two non constant polynomials. The set $K(\boldsymbol{x})$ denotes the field of rational fractions of $K[\boldsymbol{x}]$, that is the set of quotients $p(\boldsymbol{x})/q(\boldsymbol{x})$ where $p, q \in K[\boldsymbol{x}]$ and $q \neq 0$.

▶ Remark 1 (Arithmetic of $K[\boldsymbol{x}]$). We chose to use as little mathematical notion of $K[\boldsymbol{x}]$ as possible, to keep our criteria useful for people that are not familiar with multivariate polynomials. To understand the proofs, it is useful to remember that $K[\boldsymbol{x}]$ is factorial (see for instance [20, Corrolary 2.4 p 183]): any polynomial in $K[\boldsymbol{x}]$ admits a unique factorization as a product of irreducible polynomials. However, $K[\boldsymbol{x}]$ is not principal in general (even when $K = \mathbb{Q}$), nor euclidian; in particular, the Bezout identity does not hold anymore. Hence there is no canonical multivariate equivalent to the euclidian division, and similarly there is no canonical partial fraction decomposition.

Multivariate series. We write $K[[\boldsymbol{x}]]$ for the ring of formal multivariate series with (commutative) indeterminates $\boldsymbol{x}$ and coefficients in K, which is the set of infinite polynomials of the form

$$\sum_{\boldsymbol{v} \in \mathbb{N}^d} a_{\boldsymbol{v}} \boldsymbol{x}^{\boldsymbol{v}} := \sum_{v_1, \ldots, v_d \in \mathbb{N}^d} a_{v_1, \ldots, v_d} x_1^{v_1} \ldots x_d^{v_d}, \quad \text{with } a_{\boldsymbol{v}} \in K \text{ for all } \boldsymbol{v} \in \mathbb{N}^d.$$

The series $\sum_{n,m} x^n y^m$ satisfies the equation $(1-x)(1-y)S(x,y) = 1$ and hence is written $\frac{1}{(1-x)(1-y)}$. Similarly, for every monomial $\boldsymbol{x}^{\boldsymbol{v}}$, the series $\sum_{n \in \mathbb{N}} \boldsymbol{x}^{n\boldsymbol{v}}$ is written $\frac{1}{1-\boldsymbol{x}^{\boldsymbol{v}}}$; for instance, the series $\sum_{n,m} x^n y^n$ is written $\frac{1}{1-xy}$. A series $f(\boldsymbol{x})$ is called algebraic over K if there exists a non null multivariate polynomial $P(\boldsymbol{x}, Y) \in K[\boldsymbol{x}, Y]$ such that $P(\boldsymbol{x}, f(\boldsymbol{x})) = 0$.

Semilinear sets. Let $d \in \mathbb{N}$. A set $L \subseteq \mathbb{N}^d$ is called *linear* if there exists a vector $\boldsymbol{c} \in \mathbb{N}^d$, and a finite set of vectors $P = \{\boldsymbol{p_1}, \ldots, \boldsymbol{p_s}\}$, called periods, such that

$$L = \{\boldsymbol{c} + \lambda_1 \boldsymbol{p_1} + \ldots + \lambda_s \boldsymbol{p_s} \ : \ \lambda_1, \ldots, \lambda_s \in \mathbb{N}\}.$$

We will denote such a set under the condensed form $\boldsymbol{c} + P^*$. A *semilinear set* $S \subseteq \mathbb{N}^d$ is a finite union of linear sets in $\mathbb{N}^d$. The generating series of a semilinear set is defined by the multivariate series $S(\boldsymbol{x}) = \sum_{\boldsymbol{v} \in S} \boldsymbol{x}^{\boldsymbol{v}}$. In the particular case where $S = \boldsymbol{c} + P^*$ is a linear set, with $P = \{\boldsymbol{p_1}, \ldots, \boldsymbol{p_s}\}$ a set of linearly independent periods over $\mathbb{Q}$, then the decomposition of a vector $\boldsymbol{v} \in S$ under the form $\boldsymbol{v} = \boldsymbol{c} + \lambda_1 \boldsymbol{p_1} + \ldots + \lambda_s \boldsymbol{p_s}$ is unique, hence:

$$S(\boldsymbol{x}) = \sum_{\lambda_1, \ldots, \lambda_r \in \mathbb{N}^r} \boldsymbol{x}^{\boldsymbol{c} + \lambda_1 \boldsymbol{p_1} + \ldots + \lambda_s \boldsymbol{p_s}} = \boldsymbol{x}^{\boldsymbol{c}} \sum_{\lambda_1 \in \mathbb{N}} (\boldsymbol{x}^{\boldsymbol{p}_1})^{\lambda_1} \ldots \sum_{\lambda_s \in \mathbb{N}} (\boldsymbol{x}^{\boldsymbol{p}_r})^{\lambda_r} = \frac{\boldsymbol{x}^{\boldsymbol{c}}}{\prod_{\boldsymbol{p} \in P} (1 - \boldsymbol{x}^{\boldsymbol{p}})}.$$

Note that by [8, 18], it is always possible to find a representation of a semilinear S under the form $S = \biguplus_{i=1}^r (\boldsymbol{c}_i + P_i^*)$, where the union is disjoint, and the vectors are linearly independent over $\mathbb{Q}$ in each P_i. Hence the generating series of a semilinear set is rational and can be deduced from such a presentation by $S(\boldsymbol{x}) = \displaystyle\sum_{i=1}^r \frac{\boldsymbol{x}^{\boldsymbol{c}_i}}{\prod_{\boldsymbol{p} \in P_i} (1 - \boldsymbol{x}^{\boldsymbol{p}})}$.

Computing generating series of semilinear sets. In practice, for the examples of this article, we will not need a representation of S of the previous form, and can compute the generating series of such sets by hand. For instance, the generating series of $\mathbb{N}^2$ is $\sum_{n,m} x^n y^m = \frac{1}{(1-x)(1-y)}$, the generating series of $S_1 = \{(n,m) : n = m\}$ is $\sum_n x^n y^n = \frac{1}{1-xy}$, the generating series of $S_2 = \{(n,m) : n \neq m\} = \mathbb{N}^2 \setminus S_1$ is $\frac{1}{(1-x)(1-y)} - \frac{1}{1-xy}$. For a union, we can add the generating series, but we need to be careful to subtract the intersection, otherwise the vectors of the intersection would be counted twice. For instance, the generating series of $S_3 = \{(n,m,p) : n = m \text{ or } n = p\}$ is $\frac{1}{(1-xy)(1-z)} + \frac{1}{(1-yz)(1-x)} - \frac{1}{1-xyz}$.

3 Infinite ambiguity

Let L be a context-free language. For $n \in \mathbb{N}$, we recall that ℓ_n denotes the number of words in L of length n. In this section, we show how the asymptotic behaviour of ℓ_n can sometimes be sufficient to prove the inherent infinite ambiguity of L.

▶ **Definition 2** (finite degree of ambiguity). *Let $k \in \mathbb{N}$. A context-free grammar G is said to be k-ambiguous if every word $w \in \mathcal{L}(G)$ admits at most k different derivation trees. Similarly a context-free language L is k-ambiguous if it can be recognized by a k-ambiguous CFG.*

If such a finite k exists, then L is said to be of bounded ambiguity, or finitely ambiguous; otherwise, L is said to be of unbounded ambiguity, or infinitely ambiguous.

Infinitely ambiguous languages can arise from the concatenation of simple unambiguous languages; for instance, the language Pal of palindromes is unambiguous, but the language $Pal^2 = \{w_1 w_2 : w_1, w_2 \in Pal\}$ is infinitely ambiguous [7]. For infinitely ambiguous grammars, the functions $f(n)$ upper-bounding the number of different derivations of words of length n have been well studied [31, 32, 33]. Note that deciding infinite ambiguity is also undecidable [15]. The usual studies on finite or infinite ambiguity rely generally on iterations with Ogden's lemma or Ullian and Ginsburg's criteria (see for instance [29] which gives examples, for each $k \in \mathbb{N}$, of arbitrary inherently k-ambiguous on $a^*b^*c^*$). In this section we propose a novel approach based on generating series and their asymptotic behaviour.

3.1 An analytic criterion for infinite ambiguity

Let G be a context-free grammar such that every word $w \in \mathcal{L}(G)$ has a finite number of derivation. We call $G(x) = \sum_{n\in\mathbb{N}} g_n x^n$ the generating series of the derivation trees of G, where g_n denotes the number of derivation trees for words of $\mathcal{L}(G)$ of length n. Then, by the Chomsky-Schützenberger theorem [6], $G(x)$ is algebraic. More precisely, $G(x)$ belongs to a more restrictive class of algebraic series, called $\mathbb{N}$-algebraic series, for which the asymptotic behaviour of the coefficient has been well studied:

▶ **Proposition 3** (Critical exponents of $\mathbb{N}$-algebraic series [1]). *Let $G(z) = \sum_n g_n z^n$ be an $\mathbb{N}$-algebraic series. If G has a unique singularity on its circle of convergence $|z| = 1/\beta$, then*

$$g_n \sim_{n\to\infty} \frac{C}{\Gamma(1+\alpha)} n^\alpha \beta^n \,, \tag{1}$$

where C, β are non negative algebraic constants, and α belongs to the following set:

$$\mathbb{D}_2 := \{-1 - 2^{-(k+1)} : k \geq 0\} \cup \left\{-1 + \frac{r}{2^k} : k \geq 0, r \geq 1\right\} .$$

If $G(z)$ has several dominant singularities, then there exists a non negative integer p such that for every $s \in [0, p-1]$, either $g_{s+np} = 0$ for all n sufficiently large, or g_{s+np} has an asymptotic behaviour of the form of (1), where each constant depends on s.

We now derive the following criterion for infinite ambiguity, where we recall that $\mathbb{D}_2$ is defined in Proposition 3, and $n \equiv s[p]$ means that n is congruent to s modulo p:

▶ **Theorem 4.** *Suppose that it is not possible to find an integer $p \in \mathbb{N}_{>0}$ such that for all integer $s \in \{0, \dots, p-1\}$, for all $n \equiv s[p]$, $\ell_n = 0$ or ℓ_n satisfies a relation of the form $\ell_n = \Theta(\beta_s^n n^{\alpha_s})$ with $\beta_s > 0$ algebraic, and $\alpha \in \mathbb{D}_2$. Then L is infinitely ambiguous.*

Proof. We prove the contraposition: assume that L is k-ambiguous for some $k \in \mathbb{N}_{>0}$, and let us show that it is possible to find an integer $p > 0$ such that for all $s \in [0, p-1]$, for all $n \equiv s[p]$, $\ell_n = 0$ or ℓ_n satisfies a relation of the form $\ell_n = \Theta(\beta_s^n n^{\alpha_s})$ with $\beta_s > 0$ algebraic, and $\alpha \in \mathbb{D}_2$.

Let $L(x) = \sum_n \ell_n x^n$ the generating series of L, and $G(x) = \sum_n g_n x^n$ the generating series of the derivations of G. Then by definition of k-ambiguity, for every $n \in \mathbb{N}$, $\ell_n \leq g_n \leq k\ell_n$. In other words, $\frac{g_n}{k} \leq \ell_n \leq g_n$, which implies that $\ell_n = \Theta(g_n)$, where g_n is the coefficient of an $\mathbb{N}$-algebraic series. By Proposition 3, there exists a non negative integer p such that for every $s \in \{0, \dots, p-1\}$, either $g_{s+np} = 0$ for all n sufficiently large, or g_{s+np} has an asymptotic behaviour of the form of (1).

If $g_{s+np} = 0$ for all n sufficiently large, then so is ℓ_{s+np}. If $g_{s+np} \neq 0$ for n sufficiently large, then there exist C a constant, β_s a non negative algebraic number, and $\alpha_s \in \mathbb{D}_2$ such that $g_n \sim \frac{C}{\Gamma(1+\alpha_s)} n^{\alpha_s} \beta_s^n$ when $n \to \infty$ with $n \equiv s[p]$. Hence $\ell_n = \Theta(\beta_s^n n^{\alpha_s})$. ◀

▶ **Corollary 5.** *Let L be a context-free language such that, as $n \to +\infty$, $\ell_n = \Theta(\beta^n n^\alpha \log(n)^s)$. If β is not algebraic, or if $s \neq 0$, or if $\alpha \notin \mathbb{D}_2$, then L is inherently infinitely ambiguous.*

Proof. By hypothesis, there exist two constants $b_1, b_2 > 0$ such that for n large enough,

$$b_1 \beta^n n^\alpha \log(n)^s \leq \ell_n \leq b_2 \beta^n n^\alpha \log(n)^s\,.$$

In particular, for n sufficiently large, $\ell_n > 0$. Without loss of generality, we can modify its first terms, and suppose that $\ell_n > 0$ for every $n \in \mathbb{N}$. Let us prove that the hypotheses of Theorem 4 are satisfied in the case where β is not algebraic, or $s \neq 0$, or $\alpha \notin \mathbb{D}_2$.

As $\ell_n > 0$ for every $n \in \mathbb{N}$, suppose by contradiction that there exists an integer $p > 0$ such that for all $n \equiv 0[p]$, ℓ_n can be expressed as $\ell_n = \Theta(\beta_0^n n^{\alpha_0})$, with β_0 a non negative algebraic constant, and $\alpha_0 \in \mathbb{D}_2$. Hence there exists two constants $c_1, c_2 > 0$ such that, for every $n \equiv 0[p]$ sufficiently large, $c_1 \beta_0^n n^{\alpha_0} \leq \ell_n \leq c_2 \beta_0^n n^{\alpha_0}$, and combining the two inequalities:

$$0 < \frac{c_1}{b_2} \leq \left(\frac{\beta}{\beta_0}\right)^n n^{\alpha - \alpha_0} \log(n)^s \leq \frac{c_2}{b_1}\,.$$

By predominance of the growth of the exponential, if $\beta_0 \neq \beta$, the term in the middle either tends to 0 or $+\infty$ and cannot be bounded by two strictly positive constants. Hence if β is not algebraic, $\beta_0 \neq \beta$ and we obtain a contradiction, so that L is infinitely ambiguous by Theorem 4. Otherwise if β is algebraic, $\beta = \beta_0$ and for all n sufficiently large with $n \equiv 0[p]$:

$$0 < \frac{c_1}{b_2} \leq n^{\alpha - \alpha_0} \log(n)^s \leq \frac{c_2}{b_1}\,.$$

Similarly, the only way for $n^{\alpha-\alpha_0} \log(n)^s$ to be bounded by two strictly positive constants is to have both $\alpha = \alpha_0$ and $s = 0$, hence if $s \neq 0$ or $\alpha \notin \mathbb{D}_2$, we obtain a contradiction, so that L is infinitely ambiguous by Theorem 4. ◀

3.2 Application to Shamir's language

Let us illustrate the method given in the previous section on Shamir's language. Let $\Sigma = \{\#, a_1, \ldots, a_k\}$ be an alphabet of $k+1$ letters, with $k \geq 2$. We consider the extended Shamir language L_k defined by :

$$L_k = \{w \in \Sigma \,|\, w = s\#us^R v \text{ with } s, u, v \in \{a_1, \ldots, a_k\}^* \text{ and } s \neq \varepsilon\},$$

where the letter # serves only as a separator, and s^R denotes the mirror[4] of s. This language is easily recognised by the *ambiguous* context-free grammar defined by the rules $S \to AB$ and $\{A \to aAa|a\#Ba,\ B \to aB|\varepsilon \,:\, a \in \Sigma \setminus \{\#\}\}$.

For $k = 2$, the language L_2 is one of the languages showed to be infinitely ambiguous by Shamir [30], using iterations on derivations similar to Ogden's lemma (the author actually shows the finer result that most words in the language of the form $s\#w$ have as many derivation trees as there are instances of s^R in w).

We propose here an analytic proof of the infinite ambiguity of the language L_k. In the following, ℓ_n denotes the number of words of L_k of length n. The whole proof relies on the following bounds:

▶ **Proposition 6.** *There exist constants $b_1, b_2 > 0$ such that for n sufficiently large,*

$$b_1 \log_k n \leq \frac{\ell_n}{k^{n-1}} \leq b_2 \log_k n\,.$$

In other words, $\ell_n = \Theta(k^{n-1} \log_k(n))$.

Applying Corollary 5 provides an analytic proof of the infinite ambiguity of Shamir's language:

▶ **Corollary 7.** *The Shamir language L_k is infinitely ambiguous.*

▶ Remark 8. In [10], the series of a weaker version of Shamir's language is shown to have infinitely many singularities. We could wonder if the number of singularities of the generating series of a language was correlated to its degree of ambiguity. This is not the case: Flajolet [10] gave examples of 2-ambiguous languages with an infinite number of singularities; on the other hand, the language L^* with $L = \{a^n b^m c^p \,:\, n = m \text{ or } n = p\}$ has a rational generating series, hence a finite number of singularities, but is infinitely ambiguous [24, Satz 4.2.1].

4 Two simple criteria on generating series for proving the inherent ambiguity of bounded languages

In this section, we revisit Ginsburg and Ullian's criteria with generating series. We develop simple methods to prove the inherent ambiguity of bounded languages without any iteration argument. Let us fix a dimension $d \geq 1$, and Σ an alphabet[5] of cardinality more than 2.

4.1 Bounded languages and Ullian and Ginsburg's criteria

Let us fix a tuple of d words $w_1, \ldots, w_d \in \Sigma^*$, denoted by $\langle w \rangle := \langle w_1, \ldots, w_d \rangle$. We use the same notation and definition of [14]. A language L is called *bounded* with respect to $\langle w \rangle$ if $L \subseteq w_1^* \ldots w_d^*$. The fonction $f_{\langle w \rangle} : \mathbb{N}^d \to w_1^* \ldots w_d^*$ is defined by $f_{\langle w \rangle}(p_1, \ldots, p_d) = w_1^{p_1} \ldots w_d^{p_d}$

[4] If $s = s_1 s_2 \ldots s_{n-1} s_n$, then $s^R = s_n s_{n-1} \ldots s_2 s_1$.
[5] Context-free languages on an alphabet of size 1 are regular languages by Parikh theorem [27].

for every $\boldsymbol{p} \in \mathbb{N}^d$. Notice that if every w_i is a distinct letter of Σ, then the function $f_{\langle w \rangle}$ is bijective, and its inverse is the Parikh image on $w_1^* \dots w_d^*$. A bounded language L with respect to $\langle w \rangle$ is called *semilinear* if $f_{\langle w \rangle}^{-1}(L)$ is a semilinear set. By [14], every bounded context-free language is semilinear. In practice, most bounded languages are defined by giving explicitly their semilinear set $f_{\langle w \rangle}^{-1}(L)$. For instance, if $L = \{a^i b^j c^k \ : \ i = j \text{ or } j = k\}$, then $f_{\langle a,b,c \rangle}^{-1}(L) = \{(i, j, k) \in \mathbb{N}^3 \ : \ i = j \text{ or } j = k\}$.

The following definition introduces a crucial class of sets associated to bounded languages:

▶ **Definition 9** (Stratified set, [13, 14]). *A subset $X \subseteq \mathbb{N}^d$ is stratified if :*

1. *every element of X has at most two non-zero coordinates ;*
2. *it is not possible to find four integers $1 \leq i < j < k < m \leq d$ and two vectors $\boldsymbol{x}, \boldsymbol{x}' \in X$ such that $x_i x'_j x_k x'_m \neq 0$. In other words, two distinct elements of X cannot have "interlacing" nonzero coordinates.*

We sometimes say abusively that a linear set is stratified if its set of periods is stratified. Stratified sets of periods play a fundamental role in the form of the semilinear sets described by context-free grammars. In [13], Ginsburg and Ullian show that a bounded language L with respect to $a_1^* \dots a_d^*$, where $\langle a \rangle = \langle a_1, \dots, a_d \rangle$ are distinct letters, is context-free if and only if $f_{\langle a \rangle}^{-1}(L)$ is a finite union of linear sets, each with a stratified set of periods. They specialized this result for unambiguous bounded languages:

▶ **Theorem 10** (Ginsburg and Ullian criteria, [14]). *Let L be a context-free language bounded with respect to $\langle w \rangle = \langle w_1, \dots, w_d \rangle$. Then L is inherently ambiguous if and only if $f_{\langle w \rangle}^{-1}(L)$ is not a finite union of* disjoint *linear sets, each with a* stratified *set of periods whose vectors are* linearly independent*.*

▶ Remark 11. Note that it is not necessary, in order to use this criterion, to impose the decomposition of a word of L into $w_1^* \dots w_d^*$ to be unambiguous.

One direction of the equivalence can be easily understood in the case where every w_i are distinct symbols and the semilinear set associated to L is a disjoint union of linear sets with linearly independent stratified set of periods. One can easily build an unambiguous grammar recognising the language of each linear set (the non-interlacing condition makes it possible to order the vectors of the periods according to their non-zero pairs of coordinates, in a way that they are well nested). The other direction is the heart of Ginsburg and Ullian's theorem, and is based on deep arguments[6] about derivation trees.

As we mentioned it in the introduction, these criteria are powerful as they succeeded in leaving the world of grammars and derivation trees, to focus on the semilinear set behind the language. However, this characterisation of inherent ambiguity does not provide any tool to prove that a given semilinear set cannot be written as a finite union of disjoint stratified linear sets with independent periods. Hence, most proofs based on this result (see for instance [14, 16, 29]) mimicked on semilinear sets the iteration arguments that worked on derivation trees, without taking fully advantage of the fact that $\mathbb{N}^d$ and its semilinear sets are much more amenable to techniques of analysis or algebra than derivation trees.

The next sections are devoted to show how Ginsburg and Ullian's theorem actually translates nicely in the world of generating series, and thus allows to derive very simple criteria to prove the inherent ambiguity of many bounded languages.

[6] As one of the authors admits it in his book [12, p. 188], *"The proof of the necessity is extremely complicated"*.

4.2 The three variables criterion

The theorem of this section is a simple criterion to prove the inherent ambiguity of bounded languages using generating series. The proof relies on the criteria of Ginsburg and Ullian, and some arithmetic in $K[\boldsymbol{x}]$, including the unicity of the decomposition into irreducible factors. Even if we only need $K = \mathbb{Q}$ to apply the theorem to the examples of this article, we state it in the general case where K is an arbitrary field. In particular, with $K = \mathbb{F}_2$, it generalises the criterion of [21], which only deals with bounded languages on distinct letters.

▶ **Theorem 12** (Three variables criterion). *Let $L \subseteq w_1^* \dots w_d^*$ be a context-free language bounded with respect to $\langle w \rangle$. Let $S = f_{\langle w \rangle}^{-1}(L)$ its associated semilinear set, and let*

$$S(x_1, \dots, x_d) = \frac{P(x_1, \dots, x_d)}{Q(x_1, \dots, x_d)} \in K(x_1, \dots, x_d)$$

be the generating series of S, such that P and Q are polynomials of $K[x_1, \dots, x_d]$ (that need not to be coprime). Suppose that there exists an irreducible polynomial $D \in K[x_1, \dots, x_d]$ that divides Q, does not divide P, and depends on more than three variables (in other words $D \notin K[x_i, x_j]$ for all $1 \le i, j \le d$). Then L is inherently ambiguous.

Proof. Suppose that L in unambiguous. By Ginsburg et Ullian's criteria (Theorem 10), the semilinear set S can be written under the form $S = \biguplus_{i=1}^{r}(\boldsymbol{c}_i + P_i^*)$, where the union is disjoint, each P_i is stratified, and the vectors in each set of periods P_i are linearly independent.

The disjoint union as well as the independent periods mean that this is an unambiguous description of S, such that its generating series is given by:

$$\frac{P(\boldsymbol{x})}{Q(\boldsymbol{x})} = S(\boldsymbol{x}) = \sum_{i=1}^{r} \frac{\boldsymbol{x}^{\boldsymbol{c}_i}}{\prod_{\boldsymbol{p} \in P_i}(1 - \boldsymbol{x}^{\boldsymbol{p}})} = \frac{P_2(\boldsymbol{x})}{Q_2(\boldsymbol{x})}, \text{ with } Q_2(\boldsymbol{x}) = \prod_{i=1}^{r} \prod_{\boldsymbol{p} \in P_i} (1 - \boldsymbol{x}^{\boldsymbol{p}}),$$

where P_2, Q_2 are obtained by writing the sum of fractions on the same denominator. Hence $PQ_2 = P_2Q$. The irreducible polynomial D divides Q, so it divides P_2Q, hence it divides PQ_2; as D is irreducible and does not divide P, it divides Q_2.

However, as S is stratified, no period vector $\boldsymbol{p}$ in any P_i has more than two non zero coordinates. This means that Q_2 is a product of polynomials of the form $(1 - t)$ where t is a monomial with at most two variables. Each of these polynomials admits a unique factorization in irreducible polynomials, each of them having at most two variables. By the unicity of the irreducible factorization in $K[x_1, \dots, x_d]$, D cannot divide Q_2 since it is irreducible with more than three variables. Contradiction. ◀

▶ Remark 13. As seen in the preliminaries, every semilinear set can be described unambiguously [8, 18], so that it is always possible to compute its generating series.

▶ **Proposition 14.** *The following context-free languages are inherently ambiguous:*

1. $L_1 = \{a^i b^j c^k$ *with* $i = j$ *or* $j = k\}$ *and* $L_1' = \{a^i b a^j b a^k b$ *with* $i = j$ *or* $j = k\}$
2. $L_2 = \{a^i b^j c^k$ *with* $i \ne j$ *or* $j \ne k\}$ *and* $L_2' = \{a^i b a^j b a^k b$ *with* $i \ne j$ *or* $j \ne k\}$
3. $L_3 = \{a^i b^j c^k$ *with* $i = j$ *or* $j \ne k\}$ *and* $L_3' = \{a^i b a^j b a^k b$ *with* $i = j$ *or* $j \ne k\}$
4. $C := \{w_1 w_2 \,:\, w_1, w_2 \in \{a, b\}^*$ *are palindromes*$\}$

Proof. We apply Theorem 12 (with $K = \mathbb{Q}$) by exhibiting three-variables irreducible factors in the denominator of the generating series of the semilinear sets under irreducible form.

1. The generating series $S(a, b, c)$ of the semilinear set associated to L_1 is:

$$\frac{1}{(1-ab)(1-c)} + \frac{1}{(1-bc)(1-a)} - \frac{1}{1-abc} = \frac{1-3\,a^2b^2c^2+2\,a^2b^2c+2\,ab^2c^2+2\,a^2bc-ab^2c+2\,abc^2-a^2b+2\,abc-bc^2-ac}{(1-a)(1-bc)(1-c)(1-ab)(1-abc)}$$

   The polynomial $1 - abc$ in the denominator is irreducible in $\mathbb{Q}[a, b, c]$, and has three variables. Furthermore, $1 - abc$ does not divide the numerator (it can be checked with a computer algebra software, or by hand: in $\mathbb{Q}[a, b][c]$, the numerator is of degree 2 in c, so if $1 - abc$ divided it, the numerator would be of the form $(1 - abc)(\lambda c + \mu)$ with $\lambda, \mu \in \mathbb{Q}[a, b]$, so that each monomial in c^2 in the numerator should have ab in factor, which is not the case of the monomial $-bc^2$). Hence L_1 is inherently ambiguous by Theorem 12. Notice that the generating series of the semilinear set associated to L_1' is simply $b_1b_2b_3S(a_1, a_2, a_3)$ where b_1, b_2, b_3 are associated to the three letters b, and a_1, a_2, a_3 are associated to the groups of $a's$. Hence L_1' is also inherently ambiguous.
2. The associated generating series is $\frac{1}{(1-a)(1-b)(1-c)} - \frac{1}{1-abc} = \frac{a+b+c-ab-ac-bc}{(1-a)(1-b)(1-c)(1-abc)}$. The irreducible polynomial $1 - abc$ has three variables, and does not divide the numerator, since its total degree is 3, whereas the numerator is of total degree 2. Hence L_2, and similarly L_2' are inherently ambiguous.

   ▶ Remark 15. The languages L_1 and L_2 were already proved to be inherently ambiguous in [21] with the same argument. Our criterion makes it possible to extend the criterion on word-bounded languages, to prove that L_1' and L_2' are also inherently ambiguous.
3. The generating series of the semilinear set associated to L_3 is

$$\frac{1}{(1-a)(1-b)(1-c)} - \left(\frac{1}{(1-a)(1-bc)} - \frac{1}{1-abc}\right) = \frac{3\,ab^2c^2-2\,ab^2c-2\,abc^2-b^2c^2+b^2c+bc^2+ab+ac-2\,bc-a+1}{(1-a)(1-b)(1-c)(1-bc)(1-abc)}$$

   and the proof is similar as before for both L_3 and L_3'.
4. This example illustrates why criteria on bounded languages on words are more useful than on distinct letters. The language C is known to be infinitely ambiguous [7]. Let us propose a new elementary proof of just its inherent ambiguity. Suppose that C is unambiguous. Then $\tilde{C} := C \cap ba^+ba^+abbaa^+ba^+b$ would be unambiguous, by stability of unambiguous context-free languages under intersection with a regular language [14]. As

$$\tilde{C} = \{ba^nba^mbba^pba^qb : (n = q \wedge m = p) \text{ or } (n = m \wedge p = q), n, m, p, q \in \mathbb{N}_{>0}\}$$

   is bounded with respect to $\langle b, a, b, a, b, a, b, a, b\rangle$, we associate the variables x, y, z, t to the $a's$, and u_i for $i = 1 \dots 5$ for the five b's. The generating series associated to $S' = \{(n, m, p, q) \in \mathbb{N}^4_{>0} : (n = q \wedge m = p) \,\text{or}\, (n = m \wedge p = q)\}$ is $S'(x, y, z, t) = xyzt(\frac{1}{(1-xt)(1-yz)} + \frac{1}{(1-xy)(1-zt)} - \frac{1}{1-xyzt})$. Then the generating series associated to $\tilde{C}$ is:

$$S(u_1, x, u_2, y, u_3, z, u_4, t, u_5) = u_1u_2u_3^2u_4u_5xyzt\frac{-3\,x^2z^2t^2y^2+2\,x^2zt^2y+2\,xz^2t^2y+2\,y^2tx^2z+2\,y^2txz^2\ldots}{(1-xt)(1-yz)(1-xy)(1-zt)(1-xyzt)}$$

   where we truncated the numerator due to lack of space. We can verify that the irreducible 4-variables polynomial $1 - xyzt$ does not divide the numerator, which proves that $\tilde{C}$ is inherently ambiguous. Contradiction. So C in inherently ambiguous. ◀

▶ Remark 16. To check if a polynomial of the form $\pi = 1 - \boldsymbol{x}^{\boldsymbol{v}}$ with $\boldsymbol{v} \in (\mathbb{N}_{>0})^d$ does not divide the numerator P, we could also have introduced $d - 1$ new variables y_i, and perform the substitution $x_1 \leftarrow y_1^{-v_2}, x_d \leftarrow y_{d-1}^{v_{d-1}}$ and for $1 < i < d$, $x_i \leftarrow y_{i-1}^{v_{i-1}}y_i^{-v_{i+1}}$. After this substitution, π vanishes, so if after the substitution P is not the null fraction, then it is not divisible by π. This chained substitution aims specifically at cancelling π, and it is not difficult to show that if $\pi' = 1 - \boldsymbol{x}^{\boldsymbol{v}'}$ vanishes after the substitution, then $\boldsymbol{v}'$ and $\boldsymbol{v}$ are linearly dependent over $\mathbb{Q}$. This trick will be used with $d = 2$ for the second criterion of this article.

The last language of the previous proposition shows that our criteria can also be useful to prove the inherent ambiguity of non bounded languages. Here we give an other example. The language of primitive words $\mathcal{P}$, defined formally by $\mathcal{P} = \{w \in \Sigma^* \mid \forall u \in \Sigma^*, w \in u^* \Rightarrow u = w\}$, is the set of words that are not the power of a smaller word. This language is challenging, as it is still an open question to know if it is context-free. In 1994, [28] showed that the generating series of $\mathcal{P}$ is not algebraic, and hence that if it was context-free, then it would be inherently ambiguous. We propose a new proof of this fact.

▶ **Proposition 17** ([28]). *The language of primitive words in not an unambiguous context-free language.*

Proof. The language $\mathcal{P} \cap a^*ba^*ba^*b = \{a^nba^mba^pb : n \neq m \text{ or } m \neq p\} = L_2'$ is inherently ambiguous by Proposition 14. ◀

Related work A special case of Theorem 12 has already been proved by [21], in the case where each w_i is a distinct letter and $K = \mathbb{F}_2$, using completely different techniques: the author focused on $GF(2)$ grammars, a class of context-free grammars for which union is replaced by symmetric difference, and the concatenation of two languages K and L is replaced by a special concatenation $K \odot L$ which keeps only the words w of $K \cdot L$ which admit an odd number of decompositions of the form $w = w_k w_\ell$ with $w_k \in K$ and $w_\ell \in L$. In [21], the author studies the generating series associated to bounded languages in $a_1^* \dots a_d^*$ recognized by a $GF(2)$ grammar, and shows that the irreducible polynomials at their denominator can only have at most two variables. The author proves with this criterion the inherent ambiguity of the language $\{a^ib^jc^k$ with $i \neq j$ or $j \neq k\}$. At the end of the article, the author mentions Ginsburg and Ullian's criteria, saying that it would be possible to use them to prove the inherent ambiguity of the language L, but explains that the proof would not be simpler. We showed in this section that the equivalence of Ginsburg and Ullian actually translates directly into the criterion found by [21], while generalising it to bounded languages on words.

4.3 The interlacing criterion

The three variables criterion of Theorem 12 does not exploit the non interlacing condition of a stratified set. In particular, it fails on the language $L = \{a^nb^ma^pb^q \mid n = p \text{ or } m = q\}$, as the denominator of the series of its semilinear set is $(1-ac)(1-bd)(1-a)(1-b)(1-c)(1-d)$, which only contains irreducible polynomials of at most two variables. But $(1-ac)(1-bd)$ presents two irreducible polynomials with interlaced variables, hence it is natural to wonder if this could be a sign of inherent ambiguity. If so, we need however additional conditions, as such a pattern can also occur in unambiguous languages, such as in the language $\{a^nc^n : n \geq 0\} \cup \{b^nd^n : n \geq 0\}$ whose associated series is $\frac{1}{1-ac} + \frac{1}{1-bd} = \frac{2-ac-bd}{(1-ac)(1-bd)}$.

In this section, we establish a second criterion dealing with the interlacing condition (Theorem 21). We will use several technical lemmas: Lemmas 18 and 19 are classical algebra lemmas on polynomials, while Lemma 20 studies precisely the shape of irreducible polynomials dividing the denominators of series associated to stratified linear sets.

▶ **Lemma 18** (Irreducibility of $1 - x^ny^m$). *Let $n, m \in \mathbb{N}$. The polynomial $1 - x^ny^m$ is irreducible in $\mathbb{Q}[x, y]$ if and only if $n \wedge m = 1$.*

▶ **Lemma 19.** *Let $n, m \in \mathbb{N}_{>0}$. Then $1 - x^ny^m = (1 - x^\alpha y^\beta)P(x, y)$ where $\alpha \wedge \beta = 1$, and $P(x, y)$ is a non zero polynomial whose coefficients are in $\{0, 1\}$. Furthermore $\alpha = n/(n \wedge m)$ and $\beta = m/(n \wedge m)$.*

▶ **Lemma 20.** *Let $S = \boldsymbol{c} + P^*$ a stratified linear set with linearly independent periods. Let $k \geq 1$, $n, m \geq 1$ be three integers such that $n \wedge m = 1$, and $i \neq j$ be two indices of variables, and y a fresh new variable. Then:*

- *if $(1 - x_i^n x_j^m)^k \mid \prod_{\boldsymbol{p}\in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})$, then $k = 1$;*
- *if $(1 - x_i^n x_j^m) \nmid \prod_{\boldsymbol{p}\in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})$, then $\prod_{\boldsymbol{p}\in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})|_{x_i=y^m,x_j=y^{-n}} \neq 0$, seen as en element of $\mathbb{Q}(y)[\boldsymbol{x}]$, the ring of polynomials over the field $\mathbb{Q}(y)$.*

The following theorem is our second criterion for proving the inherent ambiguity of bounded languages using the non-interlacing condition.

▶ **Theorem 21** (Interlacing criterion). *Let $L \subseteq w_1^* \dots w_d^*$ a context-free language bounded with respect to $\langle w\rangle$. Let us denote by $S = f_{\langle w\rangle}^{-1}(L)$ its semilinear set, and $S(x_1, \dots, x_d) = \frac{P(x_1,\dots,x_d)}{Q(x_1,\dots,x_d)} \in \mathbb{Q}[x_1, \dots, x_d]$ its generating series, with P and Q two polynomials, non necessarily coprime. Suppose that:*

1. *Q is divided by two non-univariate irreducible polynomials $D(x_j, x_\ell)$ and $\pi(x_i, x_k)$ with interlaced indices $j < \ell$ and $i < k$ (i.e. $i < j < k < \ell$ or $j < i < \ell < k$);*
2. *$\pi(x_i, x_k)$ is of the form $\pi(x_i, x_k) = (1 - x_i^n x_k^m)$, with $n, m \geq 1$ and $n \wedge m = 1$;*
3. *finally, $D \nmid P|_{x_i=y^m,x_k=y^{-n}}$ in $\mathbb{Q}(y)[\boldsymbol{x}]$, where y is a fresh new variable.*

Then L is inherently ambiguous.

Proof. Toward a contradiction, suppose that L is unambiguous. By Theorem 10, S can be written under the form $S = \biguplus_{s=1}^r (\boldsymbol{c}_s + P_s^*)$, where the union is disjoint, the periods P_i are stratified, and the vectors in each P_i are linearly independent. Its generating series is then:

$$\frac{P(\boldsymbol{x})}{Q(\boldsymbol{x})} = S(\boldsymbol{x}) = \sum_{s=1}^{r} \frac{\boldsymbol{x}^{\boldsymbol{c}_s}}{\prod_{\boldsymbol{p}\in P_s}(1 - \boldsymbol{x}^{\boldsymbol{p}})}$$

By hypothesis, $P|_{x_i=y^m,x_k=y^{-n}} \neq 0$ (as D always divides 0), so $\pi(x_i, x_k)$ does not divide P, and D does not divide P (otherwise D would divide $P|_{x_i=y^m,x_k=y^{-n}}$ as it is not affected by the substitution). Hence both π and D are irreducible polynomials of Q, that stay in the denominator after writing the fraction $S(\boldsymbol{x})$ under irreducible form. Hence they divide the least common multiple of every $\prod_{\boldsymbol{p}\in P_s}(1 - \boldsymbol{x}^{\boldsymbol{p}})$. Let us write $Q = (1 - x_i^n x_k^m)D(x_j, x_\ell)\tilde{Q}(\boldsymbol{x})$. Note that by Lemma 20, no irreducible factor of $\tilde{Q}(\boldsymbol{x})$ that stays after writing P/Q under irreducible form cancels at $x_i = y^m, x_k = y^{-n}$. Hence if $\tilde{Q}(\boldsymbol{x})|_{x_i=y^m,x_k=y^{-n}} = 0$, this means that an irreducible factor common between $\tilde{Q}$ and P cancels with the substitution, but this is not possible since $P|_{x_i=y^m,x_k=y^{-n}} \neq 0$. So $\tilde{Q}(\boldsymbol{x})|_{x_i=y^m,x_k=y^{-n}} \neq 0$.

Let us write I_1 the set of indices s such that $(1 - x_i^n x_k^m) \mid \prod_{\boldsymbol{p}\in P_s}(1 - \boldsymbol{x}^{\boldsymbol{p}})$, and I_2 its complement. For every $s \in I_1$, let us write $\prod_{\boldsymbol{p}\in P_s}(1 - \boldsymbol{x}^{\boldsymbol{p}}) = (1 - x_i^n x_k^m)R_s(\boldsymbol{x})$. By Lemma 20, $R_s|_{x_i=y^m,x_k=y^{-n}} \neq 0$, and by the non interlacing condition, no irreducible factor of R_s is a polynomial in exactly both variables x_j, x_ℓ. Hence, no irreducible factor[7] of $R_s|_{x_i=y^m,x_k=y^{-n}}$ in $\mathbb{Q}(y)[\boldsymbol{x}]$ is a polynomial in exactly both variables x_j and x_ℓ.

By multiplying everything by π, we obtain the following equality in $\mathbb{Q}(\boldsymbol{x})$:

$$\sum_{s\in I_1} \frac{\boldsymbol{x}^{\boldsymbol{c}_s}}{R_s(\boldsymbol{x})} + (1 - x_i^n x_k^m) \sum_{s\in I_2} \frac{\boldsymbol{x}^{\boldsymbol{c}_s}}{\prod_{\boldsymbol{p}\in P_s}(1 - \boldsymbol{x}^{\boldsymbol{p}})} = \frac{P(\boldsymbol{x})}{D(x_j, x_\ell)\tilde{Q}(\boldsymbol{x})}\,.$$

For every $s \in I_2$, $\prod_{\boldsymbol{p}\in P_s}(1 - \boldsymbol{x}^{\boldsymbol{p}})|_{x_i=y^m,x_k=y^{-n}} \neq 0$ since $\pi \nmid \prod_{\boldsymbol{p}\in P_s}(1 - \boldsymbol{x}^{\boldsymbol{p}})$, by Lemma 20. Consequently, for every $s \in I_2$, $\left.\frac{\boldsymbol{x}^{\boldsymbol{c}_s}}{\prod_{\boldsymbol{p}\in P_s}(1 - \boldsymbol{x}^{\boldsymbol{p}})}\right|_{x_i=y^m,x_k=y^{-n}}$ is a well defined rational fraction

[7] As $\mathbb{Q}(y)$ is a field, $\mathbb{Q}(y)[\boldsymbol{x}]$ is also factorial.

on $\boldsymbol{x}$ with coefficients in $\mathbb{Q}(y)$. Hence, by evaluating at $x_i = y^m, x_k = y^{-n}$, we obtain the following equality on $\mathbb{Q}(y)(\boldsymbol{x})$:

$$\sum_{s \in I_1} \frac{\boldsymbol{x}^{\mathbf{c}_s}|_{x_i=y^m,x_k=y^{-n}}}{R_s(\boldsymbol{x})|_{x_i=y^m,x_k=y^{-n}}} = \frac{P(\boldsymbol{x})|_{x_i=y^m,x_k=y^{-n}}}{D(x_j,x_\ell)\tilde{Q}(\boldsymbol{x})|_{x_i=y^m,x_k=y^{-n}}}.$$

As $D(x_j, x_\ell)$ has exactly two variables x_j and x_ℓ, it is unchanged by the substitution. Furthermore, it is easy to see that an irreducible polynomial of $\mathbb{Q}[\boldsymbol{x}]$ remains irreducible in $\mathbb{Q}(y)[\boldsymbol{x}]$. As $D \nmid P|_{x_i=y^m,x_k=y^{-n}}$, D stays an irreducible factor in $\mathbb{Q}(y)[\boldsymbol{x}]$ of the denominator of the fraction $\sum_{s\in I_1} \frac{x^{\mathbf{c}_s}|_{x_i=y^m,x_k=y^{-n}}}{R_s|_{x_i=y^m,x_k=y^{-n}}}$ once put under irreducible form.

However, none of the polynomials $R_s|_{x_i=y^m,x_k=y^{-n}}$ have irreducible factors that depend on both variables x_j, x_ℓ. We obtain a contradiction when reducing the sum on the same denominator. So L is inherently ambiguous. ◀

▶ Remark 22. The last condition can be in practice replaced by the weaker condition that there exists a rational number $\alpha \in \mathbb{Q}_{>0} \setminus \{1\}$ such that $D \nmid P|_{x_i=\alpha^m,x_k=\alpha^{-n}}$.

▶ Remark 23. If D is of the form $1 - x_j^p x_\ell^q$, by Remark 16 the last condition can be in practice replaced by the weaker condition that $P|_{x_i=y^m,x_j=z^q,x_k=y^{-n},x_\ell=z^{-p}}$ is a non-null fraction, with y, z two fresh variables.

We now use the interlacing criterion to prove the following proposition:

▶ **Proposition 24.** *The following context-free languages are inherently ambiguous:*

1. $L_1 = \{a^i b^j c^k d^\ell \;:\; i = k \text{ or } j = \ell\}$
2. $L_2 = \{a^i b^j c^k d^\ell \;:\; i \neq k \text{ or } j \neq \ell\}$
3. $L_3 = \{a^i b^j c^k d^\ell \;:\; i = k \text{ or } j \neq \ell\}$ *(and similarly* $L_4 = \{a^i b^j c^k d^\ell \;:\; i \neq k \text{ or } j = \ell\}$*)*
4. $L_2' = \{a^i b^j c^k d^\ell \;:\; 3i \neq 5k \text{ or } 2j \neq 3\ell\}$
5. $L_4 = \{a^i b^j c^k d^\ell \;:\; i < k \text{ or } i + j < k + l\}$

Proof. We illustrate in the proofs several ways of verifying the hypotheses of our criterion.

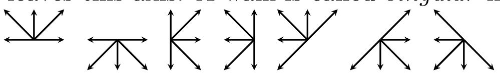
1. The generating series of the semilinear set is:
$\frac{1}{(1-ac)(1-b)(1-d)} + \frac{1}{(1-bd)(1-a)(1-c)} - \frac{1}{(1-ac)(1-bd)} = \frac{1-ab-ac-ad-bc-bd-cd+2\,abc+2\,abd+2\,acd+2\,bcd-3\,abcd}{(1-ac)(1-bd)(1-a)(1-b)(1-c)(1-d)}$
Then define $D(b,d) := 1 - bd$ and $\pi(a,c) := 1 - ac$, which are both irreducible and their variables are interlaced. Let P be the numerator $1 - ab - ac - ad - bc - bd - cd + 2\,abc + 2\,abd + 2\,acd + 2\,bcd - 3\,abcd$.
As $P|_{a=y,c=1/y} = \frac{2\,y^2bd-4\,ybd-y^2b-y^2d+2\,bd+2\,yb+2\,yd-b-d}{y}$, is of degree 1 in b, it is not divisible by $1 - bd$ in $\mathbb{Q}(y)[b,d]$. By Theorem 21, L_1 is inherently ambiguous.
2. The generating series of the semilinear set is:
$\frac{1}{(1-a)(1-b)(1-c)(1-d)} - \frac{1}{(1-ac)(1-bd)} = \frac{abc+abd+acd+bcd-ab-2\,ac-ad-bc-2\,bd-cd+a+b+c+d}{(1-ac)(1-bd)(1-a)(1-b)(1-c)(1-d)}$
Still define $\pi := 1 - ac$, $D := 1 - bd$ and P be the numerator. As $(1 - bd) \nmid P|_{a=2,c=1/2} = \frac{1}{2}(bd - b - d - 1)$, L_2 is inherently ambiguous by Theorem 21 and Remark 22.
3. The generating series of the semilinear set is[8]:
$\frac{1}{(1-a)(1-b)(1-c)(1-d)} - \frac{1}{1-bd}\left(\frac{1}{(1-a)(1-c)} - \frac{1}{(1-ac)}\right) = \frac{3\,abcd-2\,abc-abd-2\,acd-bcd+ab+ac+ad+bc-bd+cd-a-c+1}{(1-a)(1-b)(1-c)(1-d)(1-bd)(1-ac)}$
Still define $\pi = (1 - ac)$, $D = (1 - bd)$, and P be the numerator. For y, z two new variables, let us compute $P|_{a=y,b=z,c=y^{-1},d=z^{-1}} = \frac{y^2z^2-2\,y^2z-2\,yz^2+y^2+4\,yz+z^2-2\,y-2\,z+1}{yz}$ which is a non null fraction. By Remark 23 and Theorem 21, L_3 is inherently ambiguous.

[8] Because $i = k \vee j \neq \ell$ is equivalent to $\neg(\neg(i = k) \wedge j = \ell)$

4. The associated generating series is $\frac{1}{(1-a)(1-b)(1-c)(1-d)} - \frac{1}{(1-b^3d^2)(1-a^5c^3)}$, that is: $S(a,b,c,d) = \frac{a^5b^3c^3d^2-a^5c^3-b^3d^2-abcd+abc+abd+acd+bcd-ab-ac-ad-bc-bd-cd+a+b+c+d}{(1-a)(1-b)(1-c)(1-d)(1-b^3d^2)(1-a^5c^3)}$. Define $\pi = (1-a^5c^3)$ and $D = (1-b^3d^2)$, which are both irreducible with interlaced variables. Let P be the numerator. Let us choose $\alpha = 2$. As $(1-b^3d^2) \nmid P|_{a=8,c=1/32} = \frac{217}{32}(bd-b-d+1)$, L_2' is inherently ambiguous[9].
5. With a little more effort, we can check that the generating series associated to L_4 is $\frac{abcd^2-acd-bd-cd+c+d}{(1-ac)(1-ad)(1-bd)(1-d)(1-c)(1-b)}$. Still define $D = 1-bd$, $\pi = 1-ac$ and P be the numerator. Let us choose $\alpha = 2$. As $P|_{a=2,c=1/2} = bd^2 - bd - d/2 + 1/2$ is not divisible by D, by Theorem 21 and Remark 22, L_4 is inherently ambiguous. ◀

▶ Remark 25. The previous proofs are based on the form of the semilinear set, and also work for their word-variant, like $\{a^iba^jba^kba^\ell b : i \neq k \text{ or } j \neq \ell\}$.

4.4 An application to the complement of walks in the quarter plane

We consider the quarter plane $\mathbb{N}^2$ immersed in $\mathbb{Z}^2$. We represent symbolically every vector of infinite norm 1 by an arrow symbol: $\leftarrow$ represents $(-1,0)$, $\searrow$ represents $(1,-1)$, etc. The set of all these symbols $\mathcal{S} = \{\leftarrow, \swarrow, \downarrow, \searrow, \rightarrow, \nearrow, \uparrow, \nwarrow\}$ is called the set of *small steps* of $\mathbb{Z}^2$. A word in $\mathcal{S}^*$ can be represented by a walk in the plane, starting from $(0,0)$, and following the vector represented by each letter. It is confined in the quadrant if every point of the path stays in the quarter plane $\mathbb{N}^2$. For $\Sigma \subseteq \mathcal{S}$, we call W_Σ the language of words that are confined in the quarter plane. The study of such walks is an active domain of research in combinatorics (see for instance [3, 4, 5, 19, 22, 23]), as they provide a large diversity of generating series. Most of these walks are however not context-free languages, as there are two degrees of liberty. However, for every Σ, the language $\Sigma^* \setminus W_\Sigma$ of walks on Σ that leave the quadrant is context-free: a pushdown automaton non deterministically chooses one axis and accepts the word if the walk leaves this axis. A walk is called *singular* if Σ is a subset of one of the following sets[10] [22]: 

It is easy to see that singular walks are in fact unidimensional: their steps constrain the walk so that it cannot cross one of the two axes (except at $(0,0)$), so that both W_Σ and $\Sigma^* \setminus W_\Sigma$ are easily unambiguous context-free [22]. On the contrary, with the two criteria of this section, we can prove the following proposition:

▶ **Proposition 26.** *The complement of every non-singular walk on the quarter plane is an inherently ambiguous context-free language.*

5 Conclusion

In conclusion, generating series are a beautiful and useful tool to study the question of inherent ambiguity on context-free languages. It would be interesting to find other criteria to study the inherent infinite ambiguity of languages that have simple asymptotic behaviour. Can we detect the infinite ambiguity of L^*, with $L = \{a^nb^mc^p : n = m \text{ or } n = p\}$ [24, Satz 4.2.1] using generating series? One lead would be to start by proving the inherent k-ambiguity of bounded languages, for a given k. For instance, if we can show that L^k is

[9] We could also have checked that $P|_{a=y^3,b=z^2,c=y^{-5},d=z^{-3}}$ is not the null fraction.

[10] The last set is not called singular nor considered in [22], since walks with such steps leave the quadrant at the first step.

inherently $f(k)$-ambiguous with $f(k) \to_{k\to\infty} \infty$ using generating series, then we could prove that L^* is inherently infinitely ambiguous. Ginsburg and Ullian's criteria (see [14, 29]) give a characterisation of the degree of ambiguity of bounded languages: a bounded context-free language is recognized by a k-ambiguous grammar if and only if its semilinear set can be decomposed as a finite union of stratified linear set, each with independent sets of periods, such that every intersection of $l > k$ of these linear sets is empty. This implies that the generating series of the semilinear set can be expressed by inclusion-exclusion as a sum of generating series of linear stratified sets and their Hadamard's product. It looks challenging to find a pattern that can only occur in the intersection of k stratified linear sets, or equivalently in the Hadamard's product of k of their generating series.

As for inherent ambiguity of bounded languages, finding inherently ambiguous languages that are not covered by Theorems 12 and 21 would be a nice challenge to improve them. We hope that having a stronger understanding on their series would help to determinate whether inherent ambiguity is decidable or not on bounded context-free languages.

References

1 Cyril Banderier and Michael Drmota. Formulae and asymptotics for coefficients of algebraic functions. *Combinatorics, Probability and Computing*, 24(1):1–53, 2015. doi:10.1017/S0963548314000728.

2 Jason P. Bell and Shaoshi Chen. Power series with coefficients from a finite set. *J. Comb. Theory, Ser. A*, 151:241–253, 2017. doi:10.1016/j.jcta.2017.05.002.

3 Alin Bostan, Frédéric Chyzak, Mark van Hoeij, Manuel Kauers, and Lucien Pech. Hypergeometric expressions for generating functions of walks with small steps in the quarter plane. *Eur. J. Comb.*, 61:242–275, 2017. doi:10.1016/j.ejc.2016.10.010.

4 Alin Bostan and Manuel Kauers. The complete generating function for Gessel walks is algebraic. *Proc. Amer. Math. Soc.*, 138(9):3063–3078, 2010. With an Appendix by Mark van Hoeij. doi:10.1090/S0002-9939-2010-10398-2.

5 Mireille Bousquet-Mélou and Marko Petkovsek. Walks confined in a quadrant are not always D-finite. *Theor. Comput. Sci.*, 307(2):257–276, 2003. doi:10.1016/S0304-3975(03)00219-6.

6 Noam Chomsky and Marcel-Paul Schützenberger. *The Algebraic Theory of Context-Free Languages*, volume 35 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, 1963. doi:10.1016/S0049-237X(08)72023-8.

7 JP Crestin. Un langage non ambigu dont le carré est d'ambiguité non bornée. In *ICALP*, pages 377–390, 1972.

8 Samuel Eilenberg and Marcel-Paul Schützenberger. Rational sets in commutative monoids. *J. Algebra*, 13(2):173–191, 1969. doi:10.1016/0021-8693(69)90070-2.

9 Georges Elencwajg. Necessary and sufficient condition for $x^n - y^m$ to be irreducible in $[x, y]$. Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/489703 (version: 2013-09-10). URL: https://math.stackexchange.com/q/489703.

10 Philippe Flajolet. Analytic models and ambiguity of context-free languages. *Theor. Comput. Sci.*, 49(2):283–309, 1987. doi:10.1016/0304-3975(87)90011-9.

11 Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

12 Seymour Ginsburg. *The Mathematical Theory of Context-Free Languages*. McGraw-Hill, Inc., USA, 1966.

13 Seymour Ginsburg and Edwin Spanier. Semigroups, Presburger formulas, and languages. *Pac. J. Math.*, 16(2):285–296, 1966.

14 Seymour Ginsburg and Joseph Ullian. Ambiguity in context free languages. *J. ACM*, 13(1):62–89, 1966. doi:10.1145/321312.321318.

15 Sheila Greibach. A note on undecidable properties of formal languages. *Mathematical Systems Theory*, 2(1):1–6, 1968. doi:10.1007/BF01691341.

16 Thomas N. Hibbard and Joseph Ullian. The independence of inherent ambiguity from complementedness among context-free languages. *J. ACM*, 13(4):588–593, October 1966. doi:10.1145/321356.321366.

17 Juha Honkala. On parikh slender languages and power series. *Journal of Computer and System Sciences*, 52(1):185–190, 1996. doi:10.1006/jcss.1996.0014.

18 Ryuichi Ito. Every semilinear set is a finite union of disjoint linear sets. *J. Comput. Syst. Sci.*, 3(2):221–231, 1969. doi:10.1016/S0022-0000(69)80014-0.

19 Manuel Kauers and Alin Bostan. Automatic classification of restricted lattice walks. *Discrete Mathematics & Theoretical Computer Science*, 2009.

20 Serge Lang. *Algebra*, volume 211. Springer Science & Business Media, 2012.

21 Vladislav Makarov. Bounded languages described by gf(2)-grammars. In Nelma Moreira and Rogério Reis, editors, *Developments in Language Theory – 25th International Conference, DLT 2021, Porto, Portugal, August 16-20, 2021, Proceedings*, volume 12811 of *Lecture Notes in Computer Science*, pages 279–290. Springer, 2021. doi:10.1007/978-3-030-81508-0_23.

22 Marni Mishna. Classifying lattice walks restricted to the quarter plane. *Journal of Combinatorial Theory, Series A*, 116(2):460–477, 2009.

23 Marni Mishna and Andrew Rechnitzer. Two non-holonomic lattice walks in the quarter plane. *Theoretical Computer Science*, 410(38-40):3616–3630, 2009.

24 Mohamed Naji. Grad der mehrdeutigkeit kontextfreier grammatiken und sprachen. Diplomarbeit, Johann Wolfgang Goethe-Universität, 1998.

25 William Ogden. A helpful result for proving inherent ambiguity. *Mathematical systems theory*, 2(3):191–194, 1968. doi:10.1007/BF01694004.

26 Rohit J Parikh. Language generating devices. *Quarterly Progress Report*, 60:199–212, 1961.

27 Rohit J. Parikh. On context-free languages. *J. ACM*, 13(4):570–581, 1966. doi:10.1145/321356.321364.

28 Holger Petersen. The ambiguity of primitive words. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 679–690. Springer, 1994.

29 Arnold L Rosenberg. A note on ambiguity of context-free languages and presentations of semilinear sets. *Journal of the ACM (JACM)*, 17(1):44–50, 1970.

30 Eliahu Shamir. Some inherently ambiguous context-free languages. *Inf. Cont.*, 18(4):355–363, 1971. doi:10.1016/S0019-9958(71)90455-4.

31 Klaus Wich. Exponential ambiguity of context-free grammars. In *Developments In Language Theory: Foundations, Applications, and Perspectives*, pages 125–138. World Scientific, 2000.

32 Klaus Wich. Sublinear ambiguity. In Mogens Nielsen and Branislav Rovan, editors, *Mathematical Foundations of Computer Science 2000*, pages 690–698, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

33 Klaus Wich. *Ambiguity functions of context-free grammars and languages.* PhD thesis, Universität Stuttgart, 2005.

A Proofs of Section 3

A.1 Proof of Proposition 6

▶ **Proposition 6.** *There exist constants $b_1, b_2 > 0$ such that for n sufficiently large,*

$$b_1 \log_k n \le \frac{\ell_n}{k^{n-1}} \le b_2 \log_k n\,.$$

In other words, $\ell_n = \Theta(k^{n-1}\log_k(n))$.

Proof. For a given prefix $s \in \Sigma^*$, we denote ℓ_n^s the number of words of size n in L_k of the form $s\#w$, and for $r \geq 1$, $\ell_n^r = \sum_{|s|=r-1} \ell_n^s$ denotes the number of words in L_k of the form $s\#w$, of length n, and such that $|s| = r-1 \geq 1$.

Upper bound. Let us fix a word s, of length $r-1$. We recall that ℓ_n^s counts the number of words of length n of the form $s\#w$, such that w contains the factor s^R. By removing this last constraint on w, we easily obtain that $\ell_n^s \leq k^{n-r}$.

Moreover, by partitioning according to the position j in the word w where the factor s^R appears, we also deduce $\ell_n^s \leq \sum_{j=0}^{n-2r+1} k^j k^{n-2r+1-j} = (n-2r+2)k^{n-2r+1} \leq nk^{n-2r+1}$.

Hence $\ell_n^s \leq \min(k^{n-r}, nk^{n-2r+1})$.

Note that $\min(k^{n-r}, nk^{n-2r+1}) = k^{n-r}$ if $r \leq \log_k n + 1$. Finally, for all $r \geq 2$, $\ell_n^r = \sum_{|s|=r-1} \ell_n^s \leq k^{r-1} \min(k^{n-r}, nk^{n-2r+1})$, and so we have:

$$\ell_n^r \leq \begin{cases} k^{n-1} & \text{if } r < \log_k n + 1 \\ nk^{n-r} & \text{if } r \geq \log_k n + 1 \end{cases}$$

Majoring ℓ_n by the sum of all ℓ_n^r, and partitioning according to the position of r with respect to $\log_k n + 1$, we obtain:

$$\begin{aligned} \ell_n &\leq \sum_{2 \leq r < \log_k n+1} \ell_n^r + \sum_{r \geq \log_k n} \ell_n^r \leq k^{n-1}(\log_k n - 1) + nk^n \sum_{r \geq \log_k n+1} k^{-r} \\ &\leq k^{n-1}(\log_k n - 1) + nk^n k^{-\log_k n - 1} \frac{k}{k-1} \\ &= k^{n-1}(\log_k n - 1) + k^{n-1} \frac{k}{k-1} \sim_{n \to \infty} k^{n-1} \log_k n \end{aligned}$$

So there exists a constant $b_2 > 0$ such that for n large enough, $\ell_n \leq b_2 k^{n-1} \log_k n$.

Lower bound. We still look at a word of size n of L_k of the form $s\#w$, with s of size $|s| = r-1$. We split w into t consecutive blocks of size $r-1$, with $t = \lfloor \frac{n-r}{r-1} \rfloor$. Note that the last remaining block is of size $r_1 = (n-r) - (r-1)t$.

We want to lower-bound the number of words of L_k associated with a prefix s of size $r-1$ by looking only at the words w having s^R as one of their t blocks. Thus:

$$\begin{aligned} \ell_n^s &\geq \operatorname{card}\{w : s^R \text{ appears in one of the } t \text{ blocks of } w, \text{ with } |w| = n-r\} \\ &= k^{n-r} - \operatorname{card}\{w : s^R \text{ does not appear in any of the } t \text{ blocks of } w, \text{ with } |w| = n-r\} \\ &= k^{n-r} - k^{r_1}(k^{r-1}-1)^t = k^{n-r}\left(1 - \left(1 - \frac{1}{k^{r-1}}\right)^t\right) \end{aligned}$$

since $r_1 - (n-r) = -t(r-1)$. As this bound depends only on $|s| = r$, we deduce that

$$\ell_n^r \geq k^{n-1}\left(1 - \left(1 - \frac{1}{k^{r-1}}\right)^t\right) .$$

Notice that $\left(1 - \frac{1}{k^{r-1}}\right)^t = \exp\left(\lfloor \frac{n-r}{r-1} \rfloor \ln\left(1 - \frac{1}{k^{r-1}}\right)\right) \leq \exp\left(-\lfloor \frac{n-r}{r-1} \rfloor \frac{1}{k^{r-1}}\right)$.

Fix r such that $r \leq 1 + \frac{\log_k n}{2}$. Then $-\frac{1}{k^{r-1}} \leq -\frac{1}{\sqrt{n}}$. Besides, for $n \geq 4$, $n - r \geq \frac{n}{2}$ and $n - \log_k n \geq \frac{n}{2}$. Hence $\left\lfloor \frac{n-r}{r-1} \right\rfloor \geq \frac{n-r}{r-1} - 1 \geq \frac{n}{\log_k n} - 1 \geq \frac{n}{2\log_k n}$. Consequently for $r \leq \frac{\log_k n}{2} + 1$:

$$1 - (1 - \tfrac{1}{k^{r-1}})^t \geq (1 - \exp(-\tfrac{\sqrt{n}}{2\log_k n})) ,$$

where the right side does not depend on r, and tends to 1 when $n \to \infty$. So for n sufficiently large, we can lower-bound it by $1/2$. Thus there exists a rank $n_0 > 0$ independent of r such that for every $n \geq n_0$ and $r \leq \frac{\log_k n}{2} + 1$, we have $\ell_n^r \geq \frac{1}{2} k^{n-1}$.

Hence, for $n \geq n_0$, $\ell_n \geq \sum_{r-1 \leq \frac{1}{2}\log_k n} \ell_n^r \geq \frac{1}{4} k^{n-1} \log_k n$. ◀

B Proofs of Section 4

B.1 Proof of Lemma 18

We need the following classical folklore lemma:

▶ **Lemma 27.** *If $f \in \mathbb{Q}[x,y]$ is homogenous, so is any of its divisors.*

Proof. Let us factorize $f = gh$, with $g, h \in \mathbb{Q}[x,y]$. We can decompose $g = \sum_{i=s}^{r} g_i$ and $h = \sum_{i=s'}^{r'} h_i$ as a sum of homogenous polynomials where for every i, h_i and g_i are either zero or of total degree i. Furthermore, let us suppose that g_s, g_r, $h_{s'}$ and $h_{r'}$ are non zero. Hence $f = (\sum_{i=s}^{r} g_i)(\sum_{i=s'}^{r'} h_i)$, and the highest total degree term of f is $g_r h_{r'}$, of total degree $r + r'$, and the lowest total degree term is $g_s h_{s'}$, of total degree $s + s'$. As f is homogenous, $r + r' = s + s'$, and as $s \leq r$ and $s' \leq r'$, we get that $s = r$ and $s' = r'$; this means that g and h are homogenous. ◀

The following lemma is also folklore, the proof is given for completeness.

▶ **Lemma 18** (Irreducibility of $1 - x^n y^m$). *Let $n, m \in \mathbb{N}$. The polynomial $1 - x^n y^m$ is irreducible in $\mathbb{Q}[x,y]$ if and only if $n \wedge m = 1$.*

Proof. If n and m are not coprime, let $\delta > 1$ be a common divisor. Then $1 - x^n y^m = 1 - (x^{n/\delta} y^{m/\delta})^\delta = (1 - x^{n/\delta} y^{m/\delta}) \sum_{k=1}^{\delta-1} x^{kn/\delta} y^{km/\delta}$ is not irreducible.

If n and m are coprime, we adapt the nice proof of [9], by making it a little more elementary and thus a little less elegant. Let us write $f = 1 - x^n y^m$, and decompose $f = gh$.

Without loss of generality, $g = (a_0 + \ldots + a_{r'} x^r y^{r'})$ with $a_0 \neq 0$, $a_{r'} \neq 0$, r' is the degree of g in the variable y, and r is the degree in x of the polynomial that is the coefficient of $y^{r'}$. Similarly, $h = (a_0^{-1} + \ldots + -a_{r'}^{-1} x^s y^{s'})$ with $a_0 \neq 0$, $a_{r'} \neq 0$, s' the degree of h in the variable y, and s the degree in x of the coefficient of $y^{s'}$. Then $r' + s' = m$, and we can suppose without loss of generality that $r' \neq 0$.

The polynomial $Y^{nm} - X^{nm} = Y^{nm} f(X^m, Y^{-n}) = Y^{nr'} g(X^m, Y^{-n}) Y^{ns'} h(X^m, Y^{-n})$ is homogenous. As $Y^{nr'} g(X^m, Y^{-n})$ and $Y^{ns'} h(X^m, Y^{-n})$ are both polynomials in $K[X,Y]$, they are homogenous by Lemma 27.

Hence $a_0 Y^{nr'} + \ldots + a_{r'} X^{mr}$ is homogenous, and $mr = nr'$. As m and n are coprime, m divides r', and as $r' \neq 0$, $m \leq r'$, and consequently $m = r'$ and $s' = 0$. So $h(x,y)$ is a polynomial $\tilde{h}(x)$ in x only, but $Y^{ns'} h(X^m, Y^{-n}) = Y^{ns'} \tilde{h}(X^m)$ is homogenous, so $\tilde{h}(X^m)$ is homogenous too. As $a_0 \neq 0$, h is a constant. So f is irreducible. ◀

B.2 Proof of Lemma 19

▶ **Lemma 19.** *Let $n, m \in \mathbb{N}_{>0}$. Then $1 - x^n y^m = (1 - x^\alpha y^\beta) P(x,y)$ where $\alpha \wedge \beta = 1$, and $P(x,y)$ is a non zero polynomial whose coefficients are in $\{0,1\}$. Furthermore $\alpha = n/(n \wedge m)$ and $\beta = m/(n \wedge m)$.*

Proof. Let us write $\delta = n \wedge m$. Then $1 - x^n y^m = (1 - x^{n/\delta} y^{m/\delta}) P(x,y)$, where $P(x,y) = \sum_{k=1}^{\delta-1} x^{kn/\delta} y^{km/\delta}$ is non zero polynomial whose coefficients are in $\{0,1\}$. By definition of gcd, $(n/\delta) \wedge (m/\delta) = 1$. ◀

B.3 Proof of Lemma 20

▶ **Lemma 20.** *Let $S = \boldsymbol{c} + P^*$ a stratified linear set with linearly independent periods. Let $k \geq 1$, $n, m \geq 1$ be three integers such that $n \wedge m = 1$, and $i \neq j$ be two indices of variables, and y a fresh new variable. Then:*

- *if* $(1 - x_i^n x_j^m)^k \mid \prod_{\boldsymbol{p} \in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})$*, then* $k = 1$*;*
- *if* $(1 - x_i^n x_j^m) \nmid \prod_{\boldsymbol{p} \in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})$*, then* $\prod_{\boldsymbol{p} \in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})|_{x_i = y^m, x_j = y^{-n}} \neq 0$*, seen as en element of* $\mathbb{Q}(y)[\boldsymbol{x}]$*, the ring of polynomials over the field* $\mathbb{Q}(y)$*.*

Proof. As the period vectors are linearly independent, there exist at most two vectors $\boldsymbol{p_1}, \boldsymbol{p_2} \in P$ such that $(1 - \boldsymbol{x}^{\boldsymbol{p_1}})$ and $(1 - \boldsymbol{x}^{\boldsymbol{p_2}})$ are in $\mathbb{Q}[x_i, x_j]$.

Let us write $(1 - \boldsymbol{x}^{\boldsymbol{p_1}}) = (1 - x_i^{n_1} x_j^{m_1}) = (1 - x_i^{n_1/d_1} x_j^{m_1/d_1)}) P_1(x_i, x_j)$, with $n_1, m_1 \geq 1$, $P_1(x_i, x_j)$ which is a non zero polynomial with coefficients in $\{0, 1\}$, and $d_1 = n_1 \wedge m_1$. In particular, $P_1(1, 1) \neq 0$. Similarly, let us write $(1 - \boldsymbol{x}^{\boldsymbol{p_2}}) = (1 - x_i^{n_2} x_j^{m_2}) = (1 - x_i^{n_2/d_2} x_j^{m_2/d_2}) P_2(x_i, x_j)$ with the same conditions and notations.

As $P_1(1, 1)$ can not be zero, P_1 is not divisible by any polynomial of the form $(1 - \boldsymbol{x}^{\boldsymbol{p}})$ (and the same holds for P_2). Furthermore, the other factors in the denominator $\prod_{\boldsymbol{p} \in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})$ are not divisible by the irreducible polynomials $(1 - x_i^{n_1/d_1} x_j^{m_1/d_1})$ et $(1 - x_i^{n_2/d_2} x_j^{m_2/d_2})$, as they do not depend on simultaneously x_i and x_j.

Finally, $(n_1/d_1, m_1/d_1) \neq (n_2/d_2, m_2/d_2)$, as otherwise we would have $d_2 \boldsymbol{p_1} = d_1 \boldsymbol{p_2}$, implying that $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ would be linearly dependent.

Hence, every irreducible polynomial of the form $(1 - x_i^n x_j^m)$ with $n \wedge m = 1$ has multiplicity at most 1 in the unique irreducible factorization of $\prod_{\boldsymbol{p} \in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})$. The first point is proved. Note that every other irreducible factor depending on both x_i and x_j are divisors of polynomials with coefficients in $\{0, 1\}$.

The second point comes from the following additional observations:

- a non null polynomial with coefficients in $\{0, 1\}$, and consequently its divisors, does not become the null fraction by replacing some of its variables by y^m or y^{-n}. Indeed, when we write the rational fraction after the substitution on irreducible form, the denominator is a power of y, and the numerator is a sum of polynomials with positive coefficients.
- for $s \notin \{i, j\}$, $1 - x_s^{p_s}$ stays the same after the substitution, while $(1 - x_i^{p_i})|_{x_i = y^m} = 1 - y^{mp_i}$ is a non null polynomial in y, and $(1 - x_j^{p_j})|_{x_j = y^{-n}} = \frac{y^{np_j} - 1}{y^{np_j}}$ is a non null element of $\mathbb{Q}(y)$.
- similarly a polynomial of the form $(1 - x_t^{p_t} x_s^{p_s})$ with $p_t, p_s \geq 1$ and $\{x_t, x_s\} \neq \{x_i, x_j\}$ does not vanish by replacing x_i by y^m and x_j by y^{-n}. For instance, for $s \notin \{i, j\}$, $(1 - x_j^{p_j} x_s^{p_s})|_{x_j = y^{-n}} = \frac{y^{np_j} - x_s^{p_s}}{y^{np_j}}$ is a non null fraction.

By the previous observations, the only irreducible factors of $\prod_{\boldsymbol{p} \in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})$ that risk canceling after the substitution $x_i = y^m, x_j = y^{-n}$ are of the form $(1 - x_i^{n_1} x_j^{m_1})$ with $n_1, n_2 \geq 1$, $n_1 \wedge n_2 = 1$ and $(n, m) \neq (n_1, n_2)$. Then, the substitution replaces such a polynomial with the fraction $(1 - y^{mn_1 - nm_1})$ in $\mathbb{Q}(y)$, which becomes null if and only if $mn_1 - nm_1 = 0$, if and only if $n = n_1$ et $m = m_1$ (as $n \wedge m = 1$ and $n_1 \wedge n_2 = 1$). Consequently, no irreducible factor of $\prod_{\boldsymbol{p} \in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})$ becomes zero in $\mathbb{Q}(y)[\boldsymbol{x}]$ after the substitution, so $\prod_{\boldsymbol{p} \in P}(1 - \boldsymbol{x}^{\boldsymbol{p}})|_{x_i = y^m, x_j = y^{-n}}$ is a product of non-null polynomials in $\mathbb{Q}(y)[\boldsymbol{x}]$, hence is non null. ◀

B.4 Computation of the last series of Proposition 24

Let us explain how we computed the series of the semilinear set associated to $L_4 = \{a^i b^j c^k d^\ell : i < k \text{ or } i + j < k + l\}$. Let us notice that $i < k$ or $i + j < k + l \Leftrightarrow \neg(i \geq k$ and $i + j \geq k + l)$.

Hence $S(a, b, c, d) = \frac{1}{(1-a)(1-b)(1-c)(1-d)} - \sum_{n \geq p \text{ and } n+m \geq p+q} a^n b^m c^p d^q$. Let us write

$$S_2(a, b, c, d) = \sum_{n \geq p \text{ and } n+m \geq p+q} a^n b^m c^p d^q = \sum_{n=0}^{+\infty} a^n \sum_{p=0}^{n} c^p \sum_{m=0}^{+\infty} b^m \sum_{q=0}^{(n-p)+m} d^q .$$

Then

$$S_2(a,b,c,d) = \frac{1}{1-d}\sum_{n=0}^{+\infty} a^n \sum_{p=0}^{n} c^p \sum_{m=0}^{+\infty} b^m (1-d^{n-p+m+1})$$

$$=\tfrac{1}{(1-b)(1-c)(1-d)} \sum_{n=0}^{+\infty} a^n (1-c^{n+1}) - \tfrac{d}{(1-d)(1-bd)} \sum_{n=0}^{+\infty} (ad)^n \sum_{p=0}^{n} (c/d)^p$$

$$=\tfrac{1}{(1-b)(1-c)(1-d)}(\tfrac{1}{1-a} - \tfrac{c}{1-ac}) - \tfrac{d}{(1-d)(1-c/d)}(\tfrac{1}{1-ad} - \tfrac{c/d}{1-ac})$$

Hence, we obtain after simplification that $S(a,b,c,d) = \frac{abcd^2-acd-bd-cd+c+d}{(1-ac)(1-ad)(1-bd)(1-d)(1-c)(1-b)}$.

B.5 Extra properties

▶ **Lemma 28** (Announced in Remark 16). *Let π be polynomial of the form $\pi = 1 - x_1^{v_1} \dots x_k^{v_k}$ with $v_1, \dots, v_d > 0$ and $v_{k+1} = \dots = v_d = 0$ (we can without loss of generality rename the variables). We introduce $k-1$ new variables y_i, and perform the substitution $x_1 \leftarrow y_1^{-v_2}, x_d \leftarrow y_{k-1}^{v_{k-1}}$ and for $1 < i < k$, $x_i \leftarrow y_{i-1}^{v_{i-1}} y_i^{-v_{i+1}}$. Notice that after this substitution, π vanishes. Suppose that a polynomial of the form $\pi' = 1 - \boldsymbol{x}^{\boldsymbol{v}'}$ vanishes after the substitution. Then $\boldsymbol{v}'$ and $\boldsymbol{v}$ are linearly dependent over $\mathbb{Q}$.*

Proof. It is easy to see that in the case where π' has a non null degree in x_i for $i > k$, then it does not vanish after the substitution. Hence π' only depends on $x_1, \dots, x_k$, and $v'_{k+1} = \dots = v'_d = 0$. After performing the substitution on $\boldsymbol{x}^{\boldsymbol{v}'}$, we hence obtain:

$$F(\boldsymbol{y}) = \left(\frac{1}{y_1^{v_2}}\right)^{v'_1} \left(\frac{y_1^{v_1}}{y_2^{v_3}}\right)^{v'_2} \dots \left(\frac{y_{k-2}^{v_{k-2}}}{y_{k-1}^{v_k}}\right)^{v'_{k-1}} \left(y_{k-1}^{v_{k-1}}\right)^{v'_k}$$

For π' to be zero, the valuation of every variable y_i must be zero in F. For $1 \leq i \leq k-1$, we notice that the valuation of y_i is equal to $v_i v'_{i+1} - v_{i+1} v'_i$, hence using a determinant notation, $\begin{vmatrix} v_i & v_{i+1} \\ v'_i & v'_{i+1} \end{vmatrix} = 0$. This means that for all $1 \leq i \leq k-1$, there exists λ_i such that $\begin{pmatrix} v_i \\ v'_i \end{pmatrix} = \lambda_i \begin{pmatrix} v_{i+1} \\ v'_{i+1} \end{pmatrix}$, with $\lambda_i \neq 0$ since all the v_i's are non null. Hence every vector $\begin{pmatrix} v_i \\ v'_i \end{pmatrix}$ is colinear to $\begin{pmatrix} v_1 \\ v'_1 \end{pmatrix}$, hence the matrix $\begin{pmatrix} v_1 & \dots & v_d \\ v'_1 & \dots & v'_d \end{pmatrix}$ has rank 1: $\boldsymbol{v}$ and $\boldsymbol{v}'$ are linearly dependent on $\mathbb{Q}$. ◀

▶ **Lemma 29.** *If D is an irreducible polynomial of $\mathbb{Q}[\boldsymbol{x}]$, and y is a fresh new variable, then D is also an irreducible polynomial of $\mathbb{Q}(y)[\boldsymbol{x}]$.*

Proof. By contradiction suppose that $D = fg$, with f, g two non constant polynomials of $\mathbb{Q}(y)[\boldsymbol{x}]$. Each coefficient of f is a rational fraction of y, of the form $p(y)/q(y)$, for which both p and q has a finite set of roots in $\mathbb{Q}$ – and the same holds for g. Hence we can find a rational number $\alpha \in \mathbb{Q}$ that is not among these roots; when evaluating the equality $D = fg$ at $y = \alpha$, then D stays the same, and f and g becomes non constant polynomials in $\mathbb{Q}[\boldsymbol{x}]$, contradicting the irreducibility of D. ◀

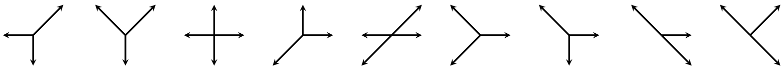
C Proof sketch of Proposition 26

▶ **Proposition 26.** *The complement of every non-singular walk on the quarter plane is an inherently ambiguous context-free language.*

Proof sketch. For $\Sigma \subseteq \mathcal{P} = \{\leftarrow, \rightarrow, \uparrow, \downarrow, \nearrow, \searrow, \nwarrow, \swarrow\}$, let us denote by $L_\Sigma = \Sigma^* \setminus W_\Sigma$ the language describing walks with steps in Σ that leave the quarter plane. Notice that if $\Sigma_1 \subseteq \Sigma_2 \subseteq \mathcal{P}$, as $L_{\Sigma_2} \cap \Sigma_1^* = L_{\Sigma_1}$, if L_{Σ_1} is inherently ambiguous, so is L_{Σ_2}.

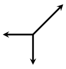
Let us denote by σ the letter-morphism representing the axial symmetry of axis $y = x$ (for instance, $\sigma(\rightarrow) = \uparrow$, $\sigma(\nearrow) = \nearrow$, and $\sigma(\nwarrow) = \searrow$). It is easy to see that for $\Sigma \subseteq \mathcal{P}$ and $w \in \Sigma^*$, $w \in L_\Sigma$ if and only if $\sigma(w) \in L_{\sigma(\Sigma)}$, so that L_Σ is inherently ambiguous if and only if $L_{\sigma(\Sigma)}$ is.

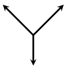
We then enumerate all non-singular sets $\Sigma \subseteq \mathcal{P}$, keep the minimal ones for inclusion, and choose arbitrarily one set per symmetry class with respect to σ. Then only nine sets remain to study:

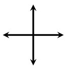


We can divide those sets in three cases. In this sketched proof, we only detail the first case.

Case $\Sigma = \{\leftarrow, \downarrow, \nearrow\}$, $\Sigma = \{\nwarrow, \downarrow, \nearrow\}$ and $\Sigma = \{\rightarrow, \uparrow, \leftarrow, \downarrow\}$

If $\Sigma =$ , notice that $L_\Sigma \cap \nearrow^* \downarrow^* \leftarrow^* = \{\nearrow^n \downarrow^m \leftarrow^p : n < m \vee n < p\}$.

If $\Sigma =$ , notice that $L_\Sigma \cap \nearrow^* \downarrow^* \nwarrow^* = \{\nearrow^n \downarrow^m \nwarrow^p : n < m \vee n < p\}$.

If $\Sigma =$ , notice that $L_\Sigma \cap (\uparrow\rightarrow)^* \downarrow^* \leftarrow^* = \{(\uparrow\rightarrow)^n \downarrow^m \leftarrow^p : n < m \vee n < p\}$.

Let us call $S = \{(n, m, p) : n < m \vee n < p\}$. As $n < m \vee n < p \Leftrightarrow \neg(n \geq m \wedge n \geq p) \Leftrightarrow \neg(n \geq m \geq p \vee n \geq p > m)$, we have:

$$\begin{aligned} S(a,b,c) &= \tfrac{1}{(1-a)(1-b)(1-c)} - \tfrac{1}{(1-abc)(1-ab)(1-a)} - \tfrac{ac}{(1-abc)(1-ac)(1-a)} \\ &= \tfrac{a^2b^2c^2 - 2\,abc - cb + b + c}{(1-ac)(1-ab)(1-abc)(1-c)(1-b)} \end{aligned}$$

We can check that $1 - abc$ does not divide the numerator. Hence by Theorem 12, L_Σ is inherently ambiguous for every set Σ of this section.

The remaining two cases are similar: we can associate to $\Sigma = \{\rightarrow, \uparrow, \swarrow\}$, $\Sigma = \{\rightarrow, \nearrow, \leftarrow, \swarrow\}$ and $\Sigma = \{\rightarrow, \nwarrow, \swarrow\}$ the semilinear set $S = \{(n, m, p) : n < p \vee m < p\}$, of generating series $\frac{(1-ab)c}{(1-c)(1-a)(1-b)(1-abc)}$; and to $\Sigma = \{\downarrow, \rightarrow, \nwarrow\}$, $\Sigma = \{\searrow, \rightarrow, \nwarrow\}$ and $\Sigma = \{\searrow, \nearrow, \nwarrow\}$ the semilinear set $S = \{(n, m, p) : n < m \vee m < p\}$, of generating series $\frac{a^2b^2c - abc - ab - bc + b + c}{(1-abc)(1-a)(1-ab)(1-b)(1-c)}$. ◀