

# Algebraic Linear Orderings

S. L. Bloom

Department of Computer Science  
Stevens Institute of Technology  
Hoboken, NJ, USA

Z. Ésik\*

Department of Informatics  
University of Szeged  
Szeged, Hungary

November 3, 2018

## Abstract

An algebraic linear ordering is a component of the initial solution of a first-order recursion scheme over the continuous categorical algebra of countable linear orderings equipped with the sum operation and the constant  $\mathbf{1}$ . Due to a general Mezei-Wright type result, algebraic linear orderings are exactly those isomorphic to the linear ordering of the leaves of an algebraic tree. Using Courcelle’s characterization of algebraic trees, we obtain the fact that a linear ordering is algebraic if and only if it can be represented as the lexicographic ordering of a deterministic context-free language. When the algebraic linear ordering is a well-ordering, its order type is an algebraic ordinal. We prove that the Hausdorff rank of any scattered algebraic linear ordering is less than  $\omega^\omega$ . It follows that the algebraic ordinals are exactly those less than  $\omega^{\omega^\omega}$ .

## 1 Introduction

Fixed points and finite systems of fixed point equations, also called recursion schemes, occur in just about all areas of computer science. For example, regular and context-free languages, rational and algebraic formal power series, finite state process behaviors can all be characterized as (components of) canonical solutions (e.g., unique, least or greatest, or initial solutions) of systems of fixed point equations. In this paper, we are interested in the solutions of systems of fixed point equations over (countable) linear orderings.

Consider the fixed point equation

$$X = \mathbf{1} + X$$

---

\*Supported in part by grant no. K 75249 from the National Foundation of Hungary for Scientific Research.

over linear orderings, where  $+$  denotes the sum operation (functor) on linear orderings, and  $\mathbf{1}$  denotes a one-point linear ordering. As it will be explained in the paper, its canonical solution is the ordinal  $\omega$  (or any linear ordering isomorphic to the ordering of the natural numbers). For another example, consider the system of fixed point equations

$$\begin{aligned} X &= Y + X \\ Y &= \mathbf{1} + Y \end{aligned}$$

The first component of its canonical solution is  $\omega^2$ , and the second component is  $\omega$ . The canonical solution of

$$X = X + \mathbf{1} + X$$

is the ordered set of the rationals.

As an example involving functors with parameters, consider the system

$$\begin{aligned} F_0 &= G(\mathbf{1}) \\ G(x) &= x + G(F(x)) \\ F(x) &= x + F(x) \end{aligned}$$

The ordinal  $\omega^\omega$  is the first component of the canonical solution of this system.

We call a linear ordering **algebraic** if it is isomorphic to the first component of the canonical solution of a system of fixed point equations of the sort

$$F_i(x_0, \dots, x_{n_i-1}) = t_i, \quad i = 1, \dots, n \quad (1)$$

where  $n_1 = 0$  and each  $t_i$  is a term composed of the function variables  $F_j$ ,  $j = 1, \dots, n$ , the individual variables  $x_0, \dots, x_{n_i-1}$ , the constant  $\mathbf{1}$  and the sum operation  $+$ . (The meaning of “canonical” will be explained below.) Moreover, we call a linear ordering **regular** if it is isomorphic to the first component of the canonical solution of a system (1) with  $n_i = 0$  for all  $i$ . Further, we call an ordinal  $\alpha$  algebraic or regular<sup>1</sup> if it is the order type of an algebraic or regular linear well-ordering.

It follows from the results in [Cour78a] and [BE10a] that up to isomorphism, an algebraic (or regular) linear ordering is isomorphic to the leaf-ordering of the frontier of an algebraic (or regular) tree. In this way, we may represent algebraic and regular linear orderings and ordinals as frontiers of algebraic or regular trees, and this is the approach we take here. Algebraic and regular trees were considered in [Cour83, Gue81].

By [Cour83], a tree is regular (algebraic, respectively) iff its “partial branch language” is a regular language (deterministic context free language (dcfl), respectively). Moreover, a “locally finite” tree is regular or algebraic iff its “branch

---

<sup>1</sup>Our regular ordinals have nothing to do with the regular ordinals of set theory.

language” is regular, or a dcfl. We will use the representation of the algebraic linear orderings as frontiers of algebraic trees to derive the result that a linear ordering is algebraic iff it is isomorphic to the lexicographic ordering of a dcfl (which additionally may be chosen to be a prefix language). A similar fact relating regular trees to regular languages also holds.

The fact that the regular ordinals are exactly those less than  $\omega^\omega$  is immediate from the results of [Heil80]. It was proved in [BE10b] that an ordinal is algebraic if and only if it is less than  $\omega^{\omega^\omega}$ , see also [BE07]. In this paper, we will prove that the Hausdorff rank of any scattered linear ordering is less than  $\omega^\omega$ , which extends one part of the result on algebraic ordinals.

We use “prefix grammars”, introduced in [BE10b], as our main tool in the analysis of algebraic scattered linear orderings represented by tree frontiers. Scattered grammars defined below, are special cases of prefix grammars. We show first that the lexicographic ordering of the context-free language generated by a scattered grammar has Hausdorff rank less than  $\omega^\omega$ . Then, we show that if the leaf ordering of an algebraic tree is scattered, it is isomorphic to the lexicographic ordering of the language generated by a scattered grammar. This shows that all scattered algebraic linear orderings are determined by scattered grammars.

This second result completes the proof that the Hausdorff rank of a scattered algebraic linear ordering is less than  $\omega^\omega$ . Since it is easy to show that all ordinals less than  $\omega^{\omega^\omega}$  are algebraic, we may conclude the result of [BE10b] that the algebraic ordinals are precisely those less than  $\omega^{\omega^\omega}$ .

The paper is organized as follows. In Section 2 we review several notions and results on linear orderings and establish a few simple facts for them that will be used in the sequel. In Section 3, we use lexicographic orderings on (prefix) languages and leaf orderings of (binary) trees to represent all countable linear orderings. In Section 4, we recall the notions of continuous categorical algebras and present two examples: the algebras of trees and linear orderings. We also introduce recursion schemes and use them to define regular and algebraic elements (or objects) of continuous categorical algebras. In particular, we obtain the notions of algebraic and regular trees and linear orderings. By a Mezei-Wright theorem, it follows that a linear ordering is algebraic iff it is isomorphic to the leaf ordering of an algebraic tree. Then in Section 5, we prove that the algebraic linear orderings are exactly those linear orderings isomorphic to the lexicographic ordering of a deterministic context-free (prefix) language. Section 6 is devoted to closure operations on algebraic linear orderings. We derive some closure properties from the closure of algebraic trees with respect to substitution, while some additional closure properties come from the representation of algebraic linear orderings as lexicographic orderings on deterministic context-free languages. In Section 7, we define prefix grammars and scattered grammars and show that the lexicographic ordering of the context-free language generated by a scattered grammar has Hausdorff rank less than  $\omega^\omega$ . Then, in

Section 8 we recall from [BE10b] a translation of recursion schemes defining algebraic trees to prefix grammars. Relying on this translation, we prove in Section 9 that the rank of every scattered algebraic linear ordering is less than  $\omega^\omega$ . Section 10 contains some concluding remarks.

## 2 Linear Orderings

For the reader's convenience, in this section we have collected together some basic definitions and results regarding linear orderings that will be used in the sequel. We start by recalling some basic definitions from [Ros82].

In this paper, a **linear ordering**  $(P, <)$  is a *countable* set  $P$  equipped with a strict linear order relation  $<$ . *Sometimes, we write just  $P$  to denote a linear ordering  $(P, <)$ .* (To force the collection of all linear orderings to be a small set, we may require that the underlying set of a linear ordering is a subset of a fixed set.) A **morphism** between linear orderings  $(P, <) \rightarrow (Q, <)$  is a function  $P \rightarrow Q$  which preserves the order relation (and is thus injective). The isomorphism class of a linear ordering is called an **order type**. Each **ordinal** may be identified with the isomorphism class of a **well-ordering**. Since linear orderings in this paper are countable, so is each ordinal. For basic facts for ordinals and ordinal arithmetic we refer to [Roit90] or [Ros82]. We will sometimes identify isomorphic linear orderings.

**Definition 2.1** *Let  $(P, <)$  be a linear ordering.*

- A **subordering** of  $P$  is a subset  $Q$  of  $P$  ordered by the restriction of  $<$  to  $Q$ . An **interval** of  $P$  is a subordering  $I$  such that for all  $x < y < z$  if  $x, z \in I$  then  $y \in I$ .
- $P$  is **dense** if  $P$  contains at least two points and whenever  $x < y$ , there is some  $z$  with  $x < z < y$ . (Note: in [Ros82], linear ordering containing zero or one point are also called dense.)
- $P$  is **scattered** if  $P$  has no subordering that is dense.

It is clear that every well-ordering is scattered. If  $(Q_x, <_x)$  is a linear ordering, for each point  $x$  in the linear ordering  $P$ , then the  **$P$ -sum** (“generalized sum”, in the terminology of [Ros82]), written

$$\sum_{x \in P} (Q_x, <_x),$$

is the set  $Q = \bigcup_{x \in P} Q_x \times \{x\}$  ordered as follows:

$$(y, x) < (y', x') \iff x < x' \text{ or } (x = x' \text{ and } y < y').$$

We mention two special cases. When  $P$  is a finite linear ordering, say  $([n], <)$ , where  $[n] = \{0, 1, \dots, n-1\}$ , we write  $\sum_{x \in P} (Q_x, <_x)$  as  $Q_0 + \dots + Q_{n-1}$ . And when for each  $x \in P$ ,  $Q_x$  is a fixed linear ordering  $(Q, <)$ , then  $\sum_{x \in P} (Q_x, <_x)$  is isomorphic to the cartesian **product**  $Q \times P$ . By using generalized sum and product, we can define the **geometric sum**  $\sum_{n \geq 0} P^n$ , where  $P$  is any linear ordering. For later use we also mention the **reverse** operation that takes a linear ordering  $P = (P, <)$  to the linear ordering  $P^* = (P, >)$ , where  $x > y$  iff  $y < x$ , for all  $x, y \in P$ .

It is known, cf. [Ros82], that any scattered sum of scattered linear orderings is scattered. Moreover, the reverse of a scattered linear ordering is scattered. Hausdorff has classified scattered linear orderings according to their rank. Recall that for an ordinal  $\alpha$ , the collection  $V_\alpha$  of linear orderings of Hausdorff rank at most  $\alpha$  is defined inductively as follows.  $V_0$  consists of the empty linear ordering and the one-point linear orderings. Assuming that  $V_\beta$  is defined for all ordinals  $\beta < \alpha$ ,  $V_\alpha$  is defined as the collection of all linear orderings

$$\sum_{n \in \mathbb{Z}} (Q_n, <_n)$$

where  $\mathbb{Z}$  is the linear ordering of the integers, and, for each  $n \in \mathbb{Z}$ ,  $Q_n \in \bigcup_{\beta < \alpha} V_\beta$ . (In [Ros82], the “index set  $\mathbb{Z}$ ” was allowed to be either a finite linear ordering, or  $\omega$ , the linear ordering of the nonnegative integers, or  $\omega^*$ , the negative integers, or  $\mathbb{Z}$ . Since  $V_0$  contains the empty linear ordering, the two definitions are equivalent.)

For a scattered linear ordering  $Q$ , the (Hausdorff) **rank of**  $Q$ ,  $r(Q)$ , is the least ordinal  $\alpha$  such that  $Q \in V_\alpha$ . Hausdorff proved that every scattered linear ordering has a (countable) rank (see [Ros82]). It is clear that for each ordinal  $\alpha$ , the collection of all linear orderings of rank  $\alpha$  (or rank at most  $\alpha$ ) is closed under the reverse operation.

We will need some facts about the ranks of scattered linear orderings. The first two facts are well-known, see [Ros82].

**Lemma 2.2** *If  $\alpha$  is an ordinal,*

$$r(\omega^\alpha) = \alpha.$$

**Lemma 2.3** *If  $P$  is a scattered linear ordering and  $Q$  is a subordering of  $P$ , then  $Q$  is scattered with  $r(Q) \leq r(P)$ . In particular, if  $\alpha, \beta$  are ordinals with  $\alpha < \beta$ , then  $r(\alpha) \leq r(\beta)$ .*

**Corollary 2.4** *Suppose that  $\beta = \omega^\gamma$  such that  $\gamma$  is a limit ordinal. Then, for any ordinal  $\alpha$ ,  $\alpha < \beta$  iff  $r(\alpha) < \gamma$ .*

*Proof.* If  $\alpha < \beta$  then since  $\gamma$  is a limit ordinal,  $\alpha < \omega^\delta$  for some  $\delta < \gamma$ . Thus,  $r(\alpha) \leq r(\omega^\delta) = \delta < \gamma$ . Conversely, if  $r(\alpha) < \gamma = r(\beta)$ , then it is not possible that  $\beta \leq \alpha$ , since otherwise we would have  $r(\beta) \leq r(\alpha)$ . Thus,  $\alpha < \beta$ .  $\square$

In particular,  $\alpha < \omega^{\omega^n}$  for some ordinal  $\alpha$  and positive integer  $n$  iff  $r(\alpha) < \omega^n$ , and  $\alpha < \omega^{\omega^\omega}$  iff  $r(\alpha) < \omega^\omega$ .

We can use Lemma 2.3 to establish the following fact for the rank of a finite sum of scattered linear orderings.

**Corollary 2.5** *Suppose  $A = B_1 + \dots + B_k$ , where each  $B_i$  is scattered, and suppose  $\alpha = \max\{r(B_i) : 1 \leq i \leq k\}$ . Then  $A$  is scattered and  $\alpha \leq r(A) \leq \alpha + 1$ .*

*Proof.* The fact that  $\alpha + 1$  is an upper bound is clear from the definition of rank. Moreover,  $\alpha$  is a lower bound by Lemma 2.3.  $\square$

We now turn to the rank of generalized sums and products of scattered linear orderings.

**Lemma 2.6** *Suppose that  $(P, <)$  is a scattered linear ordering of rank  $\alpha$  and for  $x \in P$ ,  $(Q_x, <_x)$  is a scattered linear ordering of rank at most  $\beta$ . Then the  $P$ -sum  $Q = \sum_{x \in P} (Q_x, <_x)$  is scattered of rank at most  $\beta + \alpha$ .*

*Proof.* For the fact that  $Q$  is scattered, see [Ros82]. In order to prove that  $r(Q) \leq \beta + \alpha$  we will argue by induction on  $\alpha$ . When  $\alpha = 0$  then  $P$  is either empty or a singleton set and our claim is obvious. Suppose that  $\alpha > 0$ . Then  $P$  is isomorphic to a generalized sum  $\sum_{i \in \mathbb{Z}} (P_i, <_i)$ , where each  $(P_i, <_i)$  is scattered of rank less than  $\alpha$ . Now for each fixed  $i \in \mathbb{Z}$ , the rank of  $Q_i = \sum_{x \in P_i} (Q_x, <_x)$  is strictly less than  $\beta + \alpha$ , by the induction hypothesis. Since  $Q$  is isomorphic to the generalized sum  $\sum_{i \in \mathbb{Z}} Q_i$ , it follows by the definition of the rank that  $r(Q) \leq \beta + \alpha$ .  $\square$

**Corollary 2.7** *If  $P$  and  $Q$  are scattered of rank  $\alpha$  and  $\beta$ , respectively, then  $Q \times P$  is scattered of rank at most  $\beta + \alpha$ .*

*Proof.* Indeed,  $Q \times P$  is isomorphic to  $\sum_{x \in P} Q_x$ , where  $Q_x = Q$ , for each  $x \in P$ .  $\square$

**Corollary 2.8** *If  $P$  is a scattered linear ordering of rank  $\alpha > 0$ , then  $\sum_{n \geq 0} P^n$  is scattered of rank at most  $\alpha \times \omega$ .*

*Proof.* For each  $n \geq 0$ ,  $r(P^n) \leq \alpha \times n < \alpha \times \omega$ . Thus, by the definition of rank,  $r(\sum_{n \geq 0} P^n) \leq \alpha \times \omega$ .  $\square$

In the sequel, we will deal with scattered linear orderings of rank  $< \omega^n$ , for some positive integer  $n$ . In the next corollary, we give a summary of the above facts for such linear orderings.

**Corollary 2.9** *For every  $n \geq 1$ , the collection of all scattered linear orderings of rank less than  $\omega^n$  is closed under sum and cartesian product. Moreover, if for each  $k \in \mathbb{Z}$ ,  $P_k$  is scattered with  $r(P_k) < \omega^n$ , then  $r(\sum_{k \in \mathbb{Z}} P_k) \leq \omega^n$ . Finally, if  $P$  is scattered with  $r(P) < \omega^n$ , then  $r(\sum_{k \geq 0} P^k) \leq \omega^n$ .*

**Corollary 2.10** *For every  $n \geq 1$ , the collection of all ordinals less than  $\omega^{\omega^n}$  is closed under sum and product. Moreover, if  $\alpha < \omega^{\omega^n}$  then  $\alpha^\omega \leq \omega^{\omega^n}$ .*

*Proof.* This follows from the previous corollary by noting that if  $\alpha$  is a nonzero ordinal that is the order type of a linear ordering  $P$ , then the order type of  $\sum_{k \geq 0} P^k$  is  $\alpha^\omega$ .  $\square$

We now give a characterization of the least collection of ordinals containing 0,1 which is closed under the sum, product and  $\omega$ -power operations.

**Proposition 2.11** *The least collection of ordinals which contains 0,1, and is closed under sum, product, and  $\omega$ -power consists of the ordinals less than  $\omega^{\omega^\omega}$ .*

*Proof.* The collection of ordinals less than  $\omega^{\omega^\omega}$  contains 0,1 and is closed under sum, product and  $\omega$ -power by Corollary 2.10.

The converse follows from the Cantor Normal Form (see, e.g., [Roit90]) for ordinals less than  $\omega^{\omega^\omega}$ .  $\square$

The rest of this section is devoted to unions of scattered linear orderings. The next result shows that the union of two scattered linear orderings is scattered.

**Lemma 2.12** *Suppose that  $P_0$  is a linear ordering, ordered by  $<$ , and  $P, Q \subseteq P_0$ . Then equipped with the restriction of the relation  $<$  to  $P$ ,  $Q$  and  $P \cup Q$ , respectively, each of  $P$ ,  $Q$  and  $P \cup Q$  is a linear ordering. If  $P$  and  $Q$  are scattered, then so is  $P \cup Q$ .*

*Proof.* It is obvious that each of  $P$ ,  $Q$  and  $P \cup Q$  is a linear ordering. We prove only the last statement. Assume to the contrary that  $P \cup Q$  is not scattered, so that  $P \cup Q$  contains an infinite dense subordering  $R$ . Since  $R = (R \cap P) \cup (R \cap Q)$ , either  $R \cap P$  or  $R \cap Q$  contains at least two elements. By symmetry, we may assume that  $R \cap P$  does. If  $R \cap P$  is dense, then  $P$  is not scattered, contradicting the assumption. If  $R \cap P$  is not dense, there are  $x_0 < y_0$  in  $R \cap P$  such that there is no  $z \in R \cap P$  between them. Then the set  $\{z \in R : x_0 < z < y_0\}$  is

a subset of  $Q$ . But this is a dense interval in  $R$ , contradicting the assumption that  $Q$  is scattered.  $\square$

The next goal in this section is to give an upper bound on the rank of  $P \cup Q$ , where  $P$  and  $Q$  are suborderings of a scattered linear ordering  $P_0$ . To formulate this result, we need a definition.

**Definition 2.13** *For each ordinal  $\alpha$ , we define a positive integer  $m_\alpha$  as follows. If  $\alpha$  is 0 or  $\alpha$  is a limit ordinal,  $m_\alpha = 1$ . If  $\alpha = \beta + 1$ ,*

$$m_\alpha = m_\beta + 1.$$

Note: if  $\alpha$  is a successor,  $\alpha = \lambda + k + 1$ , where  $\lambda$  is either 0 or a limit ordinal and  $0 \leq k < \omega$ , and  $m_\alpha = k + 2$ . Thus,  $1 \leq m_\alpha < \omega$ , for all ordinals  $\alpha$ .

Below, in Proposition 2.14, we suppose that  $P_0$  is a linear ordering, and  $P, Q \subseteq P_0$  are suborderings of  $P_0$ . We assume  $P$  and  $Q$  are scattered, so that  $P \cup Q$  is scattered also, by Lemma 2.12. Let  $r(P) = \alpha$  and  $r(Q) = \beta$ . We will prove the following upper bound on  $r(P \cup Q)$ .

**Proposition 2.14**

$$r(P \cup Q) \leq \min\{\alpha + \beta + m_\beta, \beta + \alpha + m_\alpha\}.$$

*Proof.* We will use induction on  $r(P)$  to prove the following claim: For  $P, Q$  as above, with  $r(P) = \alpha$  and  $r(Q) = \beta$ ,

$$r(P \cup Q) \leq \beta + \alpha + m_\alpha.$$

If  $\alpha = 0$ , either  $P = \emptyset$  or  $P$  is a singleton. If  $P = \emptyset$ ,  $P \cup Q = Q$ , so  $r(P \cup Q) = \beta = \beta + \alpha < \beta + \alpha + 1$ . If  $P = \{p\}$ ,

$$P \cup Q = Q^{<p} + \{p\} + Q^{>p},$$

where  $Q^{<p} = \{q \in Q : q < p\}$ ,  $Q^{>p} = \{q \in Q : q > p\}$ , so that, by Proposition 2.5,

$$\begin{aligned} r(P \cup Q) &\leq \max\{0, \beta\} + 1 \\ &= \beta + 1 \\ &= \beta + \alpha + m_\alpha. \end{aligned}$$

Now assume  $\alpha > 0$  and

$$P = \sum_{n \in \mathbb{Z}} P_n,$$



where the sets  $P_n$ ,  $n \in \mathbb{Z}$ , are pairwise disjoint and  $r(P_n) < \alpha$ . Define, for  $n \in \mathbb{Z}$ ,

$$Q_n := \{q \in Q : \exists p \in P_n (p < q) \text{ and } \exists p' \in P (q < p') \text{ and } \forall m > n (q < P_m)\}.$$

(Here,  $q < P_m$  means that  $q < p$  for all  $p \in P_m$ .) Note: if  $P_n = \emptyset$ , then  $Q_n = \emptyset$ . Moreover, if  $n < m$  and  $q \in Q_n$ ,  $q' \in Q_m$ , then  $q < q'$ . Also,

$$P \cup Q = Q_{-\infty} + \sum_{n \in \mathbb{Z}} (P_n \cup Q_n) + Q_{\infty},$$

where

$$\begin{aligned} Q_{-\infty} &:= \{q \in Q : q < P, \text{ i.e., } q < p, \text{ all } p \in P\} \\ Q_{\infty} &:= \{q \in Q : P < q\} \end{aligned}$$

Note that by Lemma 2.3,  $r(Q_{-\infty}), r(Q_{\infty}) \leq \beta$ , and  $r(Q_n) \leq \beta$ , for all  $n \in \mathbb{Z}$ . Thus, by Proposition 2.5 again,

$$r(P \cup Q) \leq \max\{\beta, r(\sum_{n \in \mathbb{Z}} P_n \cup Q_n)\} + 1. \quad (2)$$

Since  $\alpha$  is countable, there are two cases: either  $\alpha$  is a successor or  $\alpha$  is the least upper bound of an increasing  $\omega$ -sequence.

1. If  $\alpha$  is a successor, let  $\alpha = \alpha_1 + 1$ . Then  $r(P_n) \leq \alpha_1$ , for each  $n$ , and by the induction hypothesis,  $r(P_n \cup Q_n) \leq \beta + \alpha_1 + m_{\alpha_1}$ . Thus,

$$\begin{aligned} r(\sum_{n \in \mathbb{Z}} P_n \cup Q_n) &\leq \beta + \alpha_1 + m_{\alpha_1} + 1 \\ &= \beta + \alpha_1 + m_{\alpha}. \end{aligned}$$

So, by (2)

$$\begin{aligned} r(P \cup Q) &\leq \beta + \alpha_1 + m_{\alpha} + 1 \\ &= \beta + \alpha + m_{\alpha}. \end{aligned}$$

2. Suppose  $\alpha_1 < \alpha_2 < \dots$  and  $\alpha = \sup_k \alpha_k$ . Then,  $r(P_n) \leq \alpha_{k_n}$ , say, for each  $n \in \mathbb{Z}$ . By the induction hypothesis,  $r(P_n \cup Q_n) \leq \beta + \alpha_{k_n} + m_{\alpha_{k_n}} < \beta + \alpha$ . Thus,

$$r(\sum_{n \in \mathbb{Z}} P_n \cup Q_n) \leq \beta + \alpha,$$

so that by (2),

$$\begin{aligned} r(P \cup Q) &\leq \beta + \alpha + 1 \\ &= \beta + \alpha + m_{\alpha}. \end{aligned}$$

This ends the proof of the claim. Now, since the argument is symmetric, we have completed the proof of Proposition 2.14.  $\square$

Since for any positive integer  $n$ , the ordinals less than  $\omega^n$  are closed under addition, we obtain:

**Corollary 2.15** *Suppose that  $P_0$  is a linear ordering and  $P_1, \dots, P_k \subseteq P_0$ . If  $n \geq 1$  and each  $P_i$ ,  $1 \leq i \leq k$  is scattered and has rank less than  $\omega^n$ , then  $P_1 \cup \dots \cup P_k$  is scattered of rank less than  $\omega^n$ .*

### 3 Representation of Linear Orderings by Languages and Trees

In this section we recall the fact that each linear ordering is isomorphic to the lexicographic ordering of a (prefix) language over an ordered alphabet, or the binary alphabet. Since lexicographic orderings of prefix languages arise as frontiers of trees, it follows that every linear ordering can be represented as the leaf ordering of a (binary) tree.

Let the **alphabet** (i.e., finite nonempty set)  $A$  be linearly ordered. The set of **words** on  $A$ , written  $A^*$ , is equipped with two partial orders. The **prefix order**, written  $u <_p v$ , is defined by:

$$u <_p v \iff v = uw,$$

for some nonempty word  $w$ . The **strict, or branching order**, written  $u <_s v$ , is defined by:

$$u <_s v \iff u = xav, v = xbw,$$

for some words  $x, v, w \in A^*$  and letters  $a < b$  in  $A$ . The **lexicographic order** on  $A^*$  is defined by

$$u <_\ell v \iff u <_p v \text{ or } u <_s v.$$

It is easy to check that the lexicographic order is a linear order on  $A^*$ .

Recall from [HU79] that a language on  $A$  is a subset of  $A^*$ . A **prefix language** on  $A$  is a subset  $L$  of  $A^*$  such that if  $u \in L$  and  $uv \in L$  then  $v$  is the empty word, written  $\epsilon$ .

Note that on a prefix language, the lexicographic order agrees with the strict order relation.

Since all linear orderings in this paper are countable, we may restrict attention to subsets of  $\{0, 1\}^*$  ordered by the lexicographic order, where  $0 < 1$ .

**Proposition 3.1** *If  $(P, <)$  is a countable linear ordering, there is a (prefix) language  $L \subseteq \{0, 1\}^*$  such that  $(P, <)$  is isomorphic to  $(L, <_\ell)$ .*

*Proof.* This follows from the fact that, ordered by the lexicographic order, the set of words  $R$  denoted by the regular expression  $(0+11)^*01$  (or  $(00+11)^*01$ ) is isomorphic to the rational numbers, ordered as usual. Further, any (countable) linear ordering may be embedded in the rationals.  $\square$

**Remark 3.2** *A linear ordering is called **recursive** if it is isomorphic to a linear ordering  $(P, <)$  where  $P$  is a recursive subset of the set  $\mathbb{N}$  of nonnegative integers, and the order relation  $<$  is a recursive subset of  $\mathbb{N} \times \mathbb{N}$ . We show that each recursive linear ordering  $(P, <)$  is isomorphic to a lexicographic ordering  $(L, <_\ell)$  for some recursive (prefix) language  $L \subseteq \{0, 1\}^*$ . Indeed, suppose that  $P$  is a recursive subset of  $\mathbb{N}$  and  $<$  is a recursive subset of  $\mathbb{N} \times \mathbb{N}$ . When  $P$  is finite our claim is clear, so we may assume that  $P$  is infinite. Then let  $p_0, p_1, \dots$  be a recursive enumeration of the elements of  $P$  without repetition. Consider the regular language  $R$  denoted by the regular expression  $(0+11)^*01$ . As noted above,  $(R, <_\ell)$  is isomorphic to the ordering of the rationals. We define a morphism  $h : P \rightarrow R$  as follows. Suppose that for some  $n \geq 0$ ,  $h(p_0), \dots, h(p_{n-1})$  have already been defined. Then let  $h(p_n)$  be the lexicographically first element among the shortest words  $u$  of  $R - \{h(p_0), \dots, h(p_{n-1})\}$  such that for all  $i = 0, \dots, n-1$ ,  $u <_\ell h(p_i)$  if and only if  $p_n < p_i$ . Let  $L$  denote the set  $h(R)$ . It is clear by construction that  $L$  is recursive and  $(P, <)$  is isomorphic to  $(L, <_\ell)$ .*

We now turn to trees. Let  $\Sigma$  be any **ranked alphabet**, so that  $\Sigma$  is a finite alphabet of the form  $\Sigma = \bigcup_{n \geq 0} \Sigma_n$ , where  $\Sigma_n$  is the set of letters (or function symbols) of rank (or arity)  $n$ . Let  $r(\Sigma)$ , or just  $r$  denote the largest integer  $n$  such that  $\Sigma_n$  is not empty. Recall that  $[r] = \{0, \dots, r-1\}$ . Let  $V = \{v_0, v_1, \dots\}$  be a countable set of variables disjoint from  $\Sigma$ . One may define a (rooted, ordered) **tree**  $T \in T_\Sigma^\omega(V)$  as a partial function  $[r]^* \rightarrow \Sigma \cup V$  whose domain,  $\text{dom}(T)$ , is a prefix-closed subset of  $[r]^*$  such that if  $T(u) \in \Sigma_n$ , and if  $T(ui)$  is defined, then  $i \in [n]$ . Moreover, if  $T(u) \in V$ , then  $T(ui)$  is not defined, for all  $i \in [r]$ . The words in the domain of  $T$  are called the **vertices** of  $T$ . The **leaves** of  $T$  are the words  $u$  in the domain of  $T$  such that  $T(u) \in \Sigma_0 \cup V$ . The leaves of  $T$  form a prefix language on  $[r]$ , denoted  $\text{Fr}(T)$ . We let  $T_\Sigma^\omega$  denote the collection of all trees  $T \in T_\Sigma^\omega(V)$  such that  $T(u) \in \Sigma$  for all vertices  $u$ . Note that  $T_\Sigma^\omega$  contains the empty tree  $\perp$ . Each vertex  $x$  of a tree  $T \in T_\Sigma^\omega(V)$  is the root of a **subtree** of  $T$ , denoted  $T|_x$ , defined as follows:  $T|_x(u) = T(xu)$  for all  $u \in [r]^*$ . A tree  $T \in T_\Sigma^\omega(V)$  is called **finite** if its domain is finite and **complete** if it is not the empty tree and whenever  $T(u) \in \Sigma_n$  for some  $n$ , then  $u \cdot 0, \dots, u \cdot (n-1)$  are all in  $\text{dom}(T)$ . Moreover, a tree  $T$  is **locally finite** if for each vertex  $x$  the subtree  $T|_x$  contains at least one leaf. In particular,  $\perp$  is locally finite. The **frontier** or **leaf ordering** of a tree is the lexicographic ordering of its leaves.

*In the sequel, we will denote by  $\Delta$  the ranked alphabet containing 2 letters,  $+$*

and  $\mathbf{1}$ , such that  $+$  belongs to  $\Delta_2$  and  $\mathbf{1}$  belongs to  $\Delta_0$ . A tree in  $T_\Delta^\omega$  will be referred to as a **binary tree**. Below we will usually write  $t, s$  to denote finite trees and  $T, S$  to denote possibly infinite trees.

All linear orderings are isomorphic to frontiers of binary trees.

**Proposition 3.3** [Cour78, Cour83] *For any (countable) linear ordering  $P$ , there is a (locally finite) binary tree  $T$  such that  $(\text{Fr}(T), <_\ell)$ , the leaf ordering of  $T$ , is isomorphic to  $P$ .*

*Proof.* By Proposition 3.1, each linear ordering is isomorphic to the lexicographic ordering  $(L, <_\ell)$  of a prefix language  $L \subseteq \{0, 1\}^*$ . But any such linear ordering  $(L, <_\ell)$  is the leaf ordering of some locally finite binary tree.  $\square$

## 4 Continuous Categorical $\Sigma$ -algebras and Recursion Schemes

In this section we review the notion of continuous categorical algebras and recursion schemes over such algebras. Every recursion scheme has a canonical solution over any continuous categorical algebra giving rise to algebraic and regular objects, or elements. Two special cases will be of crucial importance for this paper. When the algebra is the  $\Delta$ -algebra of trees or linear orderings, we obtain the notions of algebraic and regular trees and linear orderings.

Suppose that  $\Sigma = \bigcup_{n \geq 0} \Sigma_n$  is a ranked alphabet. A **categorical  $\Sigma$ -algebra**  $\mathcal{A}$  consists of a (small) category, also denoted  $\mathcal{A}$ , together with a functor  $\sigma^{\mathcal{A}} : \mathcal{A}^n \rightarrow \mathcal{A}$ , for each letter  $\sigma \in \Sigma_n$ ,  $n \geq 0$ , called the operation induced by  $\sigma$ . A **morphism** of categorical  $\Sigma$ -algebras is a functor which preserves the operations up to natural isomorphism. (See [BE05, BE07, BE10a].)

We say that a categorical  $\Sigma$ -algebra  $\mathcal{A}$  is **continuous** if it has initial object and colimits of  $\omega$ -diagrams; moreover, the operations  $\sigma^{\mathcal{A}}$  are continuous, i.e., they preserve colimits of  $\omega$ -diagrams in each argument. **Morphisms** of continuous categorical  $\Sigma$ -algebras are continuous and preserve initial objects.

The notion of continuous categorical  $\Sigma$ -algebra generalizes the notion of continuous ordered  $\Sigma$ -algebra [GTWW77, Gue81], where the underlying category is a poset. Two examples of continuous categorical  $\Sigma$ -algebras are given below.

The set of trees  $T_\Sigma^\omega(V)$  (or  $T_\Sigma^\omega$ ) is turned into a  $\Sigma$ -algebra in the usual way, cf. [Gue81, Cour83, GTWW77]. Moreover, for any two trees  $T, S$ , we define  $T < S$  iff  $\text{dom}(T)$  is a proper subset of  $\text{dom}(S)$  and for each  $u \in \text{dom}(T)$ ,  $T(u) = S(u)$ . It is known that  $T_\Sigma^\omega(V)$  and  $T_\Sigma^\omega$  are continuous ordered  $\Sigma$ -algebras. Moreover,  $T_\Sigma^\omega(V)$  is freely generated by  $V$  in the class of all continuous categorical  $\Sigma$ -algebras: For any function  $h : V \rightarrow \mathcal{A}$  into a continuous categorical  $\Sigma$ -algebra,

there is up to isomorphism a unique morphism  $h^\#$  of categorical  $\Sigma$ -algebras extending  $h$ . In particular,  $T_\Sigma^\omega$  is an initial continuous categorical  $\Sigma$ -algebra. See [GTWW77, Gue81, BE10a] for more details.

The second example involves linear orderings. The category **Lin** of linear orderings has linear orderings as its objects and order preserving maps as morphisms, see also Section 2. It has as initial object the empty linear ordering denoted **0**. Moreover, **Lin** has colimits of all  $\omega$ -diagrams.

We have already defined the sum of any two linear orderings. The sum operation can naturally be extended to a functor  $\mathbf{Lin}^2 \rightarrow \mathbf{Lin}$ . The sum  $h_1 + h_2$  of morphisms  $h_i : P_i \rightarrow Q_i$ ,  $i = 1, 2$  is defined so that it agrees with  $h_i$  on  $P_i$ , for  $i = 1, 2$ . By letting **1** denote a singleton ordering, **Lin** becomes a continuous categorical  $\Delta$ -algebra. For more details, see [BE10a].

#### 4.1 Recursion schemes

Let  $\Sigma$  be a ranked set. Recall from Section 3 the definition of  $\Sigma$ -trees  $T_\Sigma^\omega(V)$  with variables in the set  $V = \{v_0, v_1, \dots\}$ . Each complete finite  $\Sigma$ -tree may be identified with a  $\Sigma$ -**term**, or term for short.

**Definition 4.1** *A recursion scheme over  $\Sigma$  is a sequence  $E$  of equations*

$$\begin{aligned} F_1(v_0, \dots, v_{k_1-1}) &= t_1 \\ &\vdots \\ F_n(v_0, \dots, v_{k_n-1}) &= t_n \end{aligned} \tag{3}$$

where  $t_i$  is a term over the ranked alphabet  $\Sigma \cup \mathcal{F}$  in the variables  $v_0, \dots, v_{k_i-1}$ , for  $1 \leq i \leq n$ , where  $\mathcal{F} = \{F_1, \dots, F_n\}$ . A recursion scheme is **regular** if  $k_i = 0$ , for each  $i$  with  $1 \leq i \leq n$ .

In the above definition,  $\Sigma \cup \mathcal{F}$  is the ranked alphabet whose letters are the letters in  $\Sigma$  together with the letters in  $\{F_1, \dots, F_n\}$  where each  $F_i$  is of rank  $k_i$ . The letters  $F_i$  are called function (or functor) variables.

In any continuous categorical  $\Sigma$ -algebra  $\mathcal{A}$ , any scheme  $E$  as in (3) induces a continuous endofunctor  $E^{\mathcal{A}}$  over the category

$$[\mathcal{A}^{k_1} \rightarrow \mathcal{A}] \times \dots [\mathcal{A}^{k_n} \rightarrow \mathcal{A}],$$

where  $[\mathcal{A}^k \rightarrow \mathcal{A}]$  denotes the category of all continuous functors  $\mathcal{A}^k \rightarrow \mathcal{A}$ . Since this category also has initial object and colimits of  $\omega$ -diagrams, it has an **initial fixed point**  $|E^{\mathcal{A}}|$  which is unique up to isomorphism (See [Ada74, Wand79]).

**Definition 4.2** *Suppose that  $\mathcal{A}$  is a continuous categorical  $\Sigma$ -algebra. We call a functor  $f : \mathcal{A}^m \rightarrow \mathcal{A}$ , **algebraic** if there is a recursion scheme  $E$  such that  $f$*

is isomorphic to  $|E|_1^A$ , the first component of the above initial solution. When  $m = 0$ ,  $f$  may be identified with an object of  $A$ , called an **algebraic object**, or **algebraic element**. An object (or element)  $a$  in  $\mathcal{A}$  is **regular** if there is a regular recursion scheme  $E$  such that  $a$  is isomorphic to  $|E|_1^A$ .

By applying the above notion to **Lin** or  $T_\Sigma^\omega(V)$  or  $T_\Sigma^\omega$ , we obtain the notions of algebraic and regular linear orderings, and algebraic and regular trees, respectively.

**Definition 4.3** For any ranked alphabet  $\Sigma$ , we call a tree  $T \in T_\Sigma^\omega(V)$  (or  $T \in T_\Sigma^\omega$ ) an **algebraic tree** (**regular tree**, respectively) if it is an algebraic element (regular element, respectively) of the continuous categorical  $\Sigma$ -algebra  $T_\Sigma^\omega(V)$  (or  $T_\Sigma^\omega$ ). We call a linear ordering  $(P, <)$  an **algebraic linear ordering** (**regular linear ordering**, respectively) if it is an algebraic object (regular object, respectively) of the continuous categorical  $\Delta$ -algebra **Lin**.

By a general Mezei-Wright theorem [BE10a], morphisms between continuous categorical algebras preserve algebraic and regular objects. Since  $T_\Delta^\omega$  is initial continuous categorical  $\Delta$ -algebra, up to natural isomorphism there is a unique morphism  $T_\Delta^\omega \rightarrow \mathbf{Lin}$ . This essentially unique morphism maps a tree to its leaf ordering. So we obtain:

**Proposition 4.4** [BE10a] A linear ordering is algebraic or regular if and only if it is isomorphic to the frontier of an algebraic or regular tree in  $T_\Delta^\omega$ .  $\square$

(For the case of regular trees and regular linear orderings see also [Cour78a].)

The leaf ordering of an algebraic tree over *any* ranked alphabet  $\Sigma$  is an algebraic linear ordering.

**Proposition 4.5** For any ranked alphabet  $\Sigma$  and any algebraic tree  $T \in T_\Sigma^\omega(V)$  there is an algebraic tree  $T' \in T_\Delta^\omega$  such that  $(\text{Fr}(T), <_\ell)$  and  $(\text{Fr}(T'), <_\ell)$  are isomorphic.  $\square$

For example, consider the system  $E$  of equations

$$\begin{aligned} F_1 &= \sigma_3(a, b, F_2(a)) \\ F_2(x) &= F_3(x, x) \\ F_3(x, y) &= \sigma_3(\sigma_1(a), F_3(x, F_3(x, y)), y) \end{aligned}$$

which involves a function symbol  $\sigma_3$  in  $\Sigma_3$  and a function symbol  $\sigma_1$  in  $\Sigma_1$ . The least solution consists of three trees  $(T_1, T_2, T_3)$  having vertices of out-degree 3

and 1, labeled  $\sigma_3$  and  $\sigma_1$ . We replace the system  $E$  by the system

$$\begin{aligned} F_1 &= +(\mathbf{1}, +(\mathbf{1}, F_2(\mathbf{1}))) \\ F_2(x) &= F_3(x, x) \\ F_3(x, y) &= +(\mathbf{1}, +(F_3(x, F_3(x, y)), y)) \end{aligned}$$

in which the right hand terms involve only the function variables and the one binary function symbol  $+$ , and the one constant symbol  $\mathbf{1}$ . If  $(T'_1, T'_2, T'_3)$  is the least solution of this second system,  $(\text{Fr}(T_i), <_\ell)$  is isomorphic to  $(\text{Fr}(T'_i), <_\ell)$ , for  $i = 1, 2, 3$ .

Several characterizations of algebraic and regular trees can be found in [GTWW77, Cour83, Gue81]. For characterizations of regular linear orderings we refer to [Cour78, BE04]. See also Section 5.

In Section 6, we will make use of the tree  $T_0$  in Example 4.6.

**Example 4.6** *Let  $\Sigma$  contain the binary symbol  $+$ , the unary symbol  $f$  and the constant  $\mathbf{1}$ . Consider the system*

$$\begin{aligned} F_0 &= F(\mathbf{1}) \\ F(x) &= +(x, F(f(x))) \end{aligned}$$

*Then the first component of the least solution of this system is the tree*

$$T_0 = +(\mathbf{1}, +(f(\mathbf{1}), +(f(f(\mathbf{1})), \dots, +(f^n(\mathbf{1}), \dots)))).$$

*Thus, this tree is algebraic. See also [Gue81], p. 39.*

## 4.2 Substitution on trees

We will derive some closure properties of algebraic linear orderings from the closure of algebraic trees with respect to substitution.

Suppose that for each  $\sigma \in \Sigma_n$  we are given a tree  $S_\sigma \in T_\Gamma^\omega(V_n)$ , i.e., a tree in  $T_\Gamma^\omega(V)$  such all leaves labeled in  $V$  are actually labeled in the set  $\{v_0, \dots, v_{n-1}\}$ . For each finite tree  $t \in T_\Sigma(V)$  we define a tree  $R = t[\sigma \mapsto S_\sigma]_{\sigma \in \Sigma}$ , sometimes denoted just  $t[\sigma \mapsto S_\sigma]$  by induction on the size of  $t$ . When  $t$  is the empty tree  $\perp$ , so is  $R$ . When  $t$  is  $x$ , for some  $x \in V$ , then  $R = x$ . Otherwise  $t$  is of the form  $\sigma(t_0, \dots, t_{m-1})$ , and we define

$$R = S_\sigma[v_0 \mapsto S_1, \dots, v_{m-1} \mapsto S_{m-1}]$$

where  $S_i = t_i[\sigma \mapsto S_\sigma]$ , for all  $i$ .

Suppose now that  $T$  is an infinite tree in  $T_\Sigma(V)$ . Then there is an ascending  $\omega$ -chain  $(t_n)_n$  of finite trees such that  $T = \sup_n t_n$ . We define

$$T[\sigma \mapsto S_\sigma] = \sup_n t_n[\sigma \mapsto S_\sigma].$$

The following facts are known, see [Cour83].

**Proposition 4.7** *Substitution is a continuous function*

$$T_{\Sigma}^{\omega}(V) \times \prod_n T_{\Gamma}^{\omega}(V_n)^{\Sigma_n} \rightarrow T_{\Gamma}^{\omega}(V).$$

*Further, the classes of algebraic and regular trees are closed under substitution.*

## 5 Algebraic Linear Orderings and Deterministic Context-free Languages

In this section we define context-free and deterministic context-free linear orderings using lexicographic orderings on cfl's and dcfl's, i.e., context-free and deterministic context-free languages. The main result of this section shows that a linear ordering is algebraic iff it is deterministic context-free. This result follows easily from Courcelle's characterization of algebraic trees by dcfl's [Cour78]. While every deterministic context-free linear ordering is context-free, it remains open whether there is a context-free linear ordering that is not deterministic context-free.

**Definition 5.1** *A linear ordering is **context-free** (**deterministic context-free**, respectively) if it is isomorphic to the lexicographic ordering of a cfl (dcfl, respectively) over some (ordered) alphabet  $A$  (or equivalently, over the 2-letter alphabet  $\{0, 1\}$ ).*

As the next proposition shows, we may restrict ourselves to cfl's or dcfl's which are prefix languages.

**Proposition 5.2** *A linear ordering is context-free (deterministic context-free, respectively) iff it is isomorphic to a linear ordering  $(L, <_{\ell})$  for some context-free (deterministic context-free, respectively) **prefix** language for some (ordered) alphabet  $A$  (or equivalently, over the 2-letter alphabet  $\{0, 1\}$ ).*

*Proof.* We only prove this fact for the two-letter alphabet  $\{0, 1\}$ . Let  $L \subseteq \{0, 1\}^*$  be a context free language. If  $L$  is not a prefix language, then let  $L' = L(-1)$ , a context-free language on the three-letter alphabet  $\{-1, 0, 1\}$ , ordered by  $-1 < 0 < 1$ . It is clear that  $L'$  is a context-free prefix language and  $(L', <_{\ell})$  is isomorphic to  $(L, <_{\ell})$ . To end the proof, let us introduce the following encoding of the three letter alphabet  $\{-1, 0, 1\}$ :  $h(-1) = 00$ ,  $h(0) = 01$  and  $h(1) = 10$ . Then  $h$  extends to an (injective) homomorphism  $\{-1, 0, 1\}^* \rightarrow \{0, 1\}^*$  as usual, and  $(L, <_{\ell})$  is isomorphic to  $(h(L'), <_{\ell})$ . Moreover, when  $L$  is deterministic, so is  $h(L')$ .  $\square$



Each tree can be represented in several ways by languages. Recall that a tree  $T \in T_\Sigma^\omega$  is nothing but a partial function  $[r]^* \rightarrow \Sigma$ , where  $r = r(\Sigma)$ , subject to certain properties. Defining a function  $[r]^* \rightarrow \Sigma$  amounts to specifying the languages  $T^{-1}(\sigma)$ , for all  $\sigma \in \Sigma$ , or a single language which is a combination of these languages. Below we review from [Cour78a, Cour83] a possible representation of trees by (partial) branch languages.

Given  $\Sigma$ , we introduce the (unranked) alphabet  $\bar{\Sigma} = \{(\sigma, i) : i \in [n], \sigma \in \Sigma_n, n > 0\}$ . When  $T$  is a tree in  $T_\Sigma^\omega$  and  $u \in \text{dom}(T)$ , we define  $\hat{u} \in \bar{\Sigma}^*$  by induction. First,  $\hat{\epsilon} = \epsilon$ . Next, when  $u = vi$  with  $T(v) = \sigma$  and  $i \in [r]$ , then  $\hat{u} = \hat{v}(\sigma, i)$ . It is clear that for any  $T$ , the **partial branch language**

$$\text{Pbr}(T) = \{\hat{u}T(u) : u \in \text{dom}(T)\}$$

over the alphabet  $\Sigma \cup \bar{\Sigma}$  completely describes  $T$ . We also define the **branch language** or **labeled frontier language** of  $T$  as the set

$$\text{Lfr}(T) = \{\hat{u}T(u) : T(u) \in \Sigma_0\}.$$

For the following result see [Cour83], Theorem 5.5.1.

**Theorem 5.3** *Suppose that  $T \in T_\Sigma^\omega$  is a **complete** tree. If  $T \in T_\Sigma^\omega$  is algebraic then  $\text{Pbr}(T)$  and  $\text{Lfr}(T)$  are dcfl's. Moreover, if  $\text{Pbr}(T)$  is a dcfl or when  $T$  is locally finite and  $\text{Lfr}(T)$  is a dcfl, then  $T$  is algebraic.*

In order to make the above result applicable to trees in  $T_\Sigma^\omega$  that are not necessarily complete, we need to describe a procedure for completing trees. Let  $\Omega$  be a new letter of rank 0 and let  $\Sigma_\Omega$  denote the ranked alphabet obtained by adding the letter  $\Omega$  to  $\Sigma$ . Then each tree  $T \in T_\Sigma^\omega$  has a **completion**  $T_\Omega \in T_{\Sigma_\Omega}^\omega$  defined as follows: For all  $u \in [r]^*$  where  $r = r(\Sigma)$ , if  $T(u)$  is defined then  $T_\Omega(u) = T(u)$ . If  $T(u)$  is not defined but  $u = \epsilon$  or  $u = vi$  for some  $v$  and  $i \in [n]$  such that  $T(v) \in \Sigma_n$ , then we let  $T_\Omega(u) = \Omega$ .

The next fact is clear:

**Lemma 5.4** *For any tree  $T \in T_\Sigma^\omega$ ,  $T$  is algebraic iff  $T_\Omega$  is.*

By Lemma 5.4 and Theorem 5.3 we immediately have:

**Corollary 5.5** *Suppose that  $T \in T_\Sigma^\omega$ . If  $T$  is algebraic then  $\text{Pbr}(T_\Omega)$  and  $\text{Lfr}(T_\Omega)$  are dcfl's. Moreover, if  $\text{Pbr}(T_\Omega)$  is a dcfl or if  $T_\Omega$  is locally finite and  $\text{Lfr}(T_\Omega)$  is a dcfl, then  $T$  is algebraic.*

Using the above corollary, we now establish the following (slight) generalization of Theorem 5.3.

**Proposition 5.6** *Suppose that  $T \in T_\Sigma^\omega$ . If  $T$  is algebraic then  $\text{Pbr}(T)$  and  $\text{Lfr}(T)$  are dcfl's. Moreover, if  $\text{Pbr}(T)$  is a dcfl or if  $\text{Lfr}(T)$  is a dcfl and  $T$  is locally finite, then  $T$  is algebraic.*

*Proof.* Suppose first that  $T$  is algebraic. Then  $T_\Omega$  is also algebraic and thus  $\text{Pbr}(T_\Omega)$  and  $\text{Lfr}(T_\Omega)$  are dcfl's as are  $\text{Pbr}(T) = \text{Pbr}(T_\Omega) \cap \overline{\Sigma}^* \Sigma$  and  $\text{Lfr}(T) = \text{Lfr}(T_\Omega) \cap \overline{\Sigma}^* \Sigma_0$ . Here we have used the well-known fact that the intersection of a dcfl with a regular language is a dcfl.

Suppose now that  $\text{Pbr}(T)$  is a dcfl and consider a deterministic pushdown automaton (dpda) accepting  $\text{Pbr}(T)$  with final states, see [HU79]. We modify this dpda by adding to the set of states a new state  $q_\Omega$  and a new accepting state  $s_\Omega$ , and by adding rules ensuring that whenever the dpda is able to move from a configuration  $c$  to an accepting configuration while reading a letter  $\sigma \in \Sigma_n$ ,  $n > 0$ , but at the same time  $c$  has no successor configuration with respect to the letter  $(\sigma, i)$  for some  $i$ , then the new dpda will be able to move first to state  $q_\Omega$  while reading  $(\sigma, i)$  and then to  $s_\Omega$  while reading  $\Omega$ . The resulting dpda accepts  $\text{Pbr}(T_\Omega)$ . Thus, by Theorem 5.3 and Lemma 5.4,  $T_\Omega$  and  $T$  are algebraic.

Suppose finally that  $T$  is locally finite and  $L = \text{Lfr}(T)$  is a dcfl. A similar but somewhat more involved construction works to show that  $\text{Lfr}(T_\Omega)$  is a dcfl. Indeed, there exists a dpda accepting  $L$  such that whenever from the initial configuration the dpda can reach a configuration while reading a word  $u$ , then there is some word  $v$  such that  $uv \in L$ . This holds because  $\text{Lfr}(T)$  is not empty and since the dpda obtained from the canonical LR(1) parser [AU72] has this property. Then we proceed as above. We add two states  $q_\Omega$  and  $s_\Omega$  and new transitions to the effect that if a given configuration  $c$  has a successor configuration for a letter  $(\sigma, i)$  but at the same time  $c$  has no successor configuration for  $(\sigma, j)$ , for some  $j$ , then the new dpda is able to move from  $c$  to state  $q_\Omega$  while reading  $(\sigma, j)$  and then to the state  $s_\Omega$  while reading the letter  $\Omega$ .  $\square$

We now consider trees over the ranked alphabet  $\Delta$  defined above. In this case we can identify the letter  $(+, 0)$  with 0 and the letter  $(+, 1)$  with 1, so that for any tree  $T$ ,  $\text{Pbr}(T)$  may be viewed as a subset of  $\{0, 1\}^* \Delta$  and  $\text{Lfr}(T)$  as a subset of  $\{0, 1\}^* \mathbf{1}$ . The main contribution of this section is:

**Theorem 5.7** *A linear ordering is algebraic iff it is deterministic context-free.*

*Proof.* Suppose that  $(P, <)$  is an algebraic linear ordering. Then there is an algebraic tree  $T \in T_\Delta^\omega$  such that  $(P, <)$  is isomorphic to  $(\text{Fr}(T), <_\ell)$ . But  $\text{Fr}(T)$  is the right quotient of the deterministic context-free language  $\text{Lfr}(T)$  with respect to the letter  $\mathbf{1}$  and is thus also a dcfl. (See [HU79].) We conclude that  $(P, <)$  is a deterministic context-free linear ordering.

Suppose now that  $(P, <)$  is isomorphic to  $(L, <_\ell)$  where  $L \subseteq \{0, 1\}^*$  is a dcfl. By Proposition 5.2, we may suppose that  $L$  is a prefix language. Define the

tree  $T \in T_\Delta^\omega$  by

$$T(u) = \begin{cases} + & \text{if } uv \in L \text{ for some } v \neq \epsilon \\ \mathbf{1} & \text{if } u \in L \\ \text{undefined} & \text{otherwise,} \end{cases}$$

for all  $u \in \{0, 1\}^*$ . Then  $T$  is a locally finite tree with  $\text{Fr}(T) = L$ . Since  $\text{Fr}(T)$  is a dcl, so is  $\text{Lfr}(T) = \text{Fr}(T)\mathbf{1}$ . Thus, since  $T$  is locally finite,  $T$  is algebraic proving that  $(P, <)$  is also algebraic.  $\square$

**Corollary 5.8** *A linear ordering  $(P, <)$  is algebraic iff  $P$  is empty or  $(P, <)$  is isomorphic to  $(\text{Fr}(T), <_\ell)$  for some locally finite algebraic tree  $T \in T_\Delta^\omega$ .*

**Open Problem.** Does there exist a context-free linear ordering that is not a deterministic context-free linear ordering?

## 6 Closure Properties of Algebraic Linear Orderings

In this section, we establish some closure properties of algebraic linear orderings. In particular, we will show that algebraic linear orderings are closed under finite sum and product as well as infinite geometric sum. These closure properties will readily imply that all ordinals less than  $\omega^{\omega^\omega}$  are algebraic.

We will derive the closure of algebraic linear orderings under sum, product and infinite geometric sum from the closure of algebraic trees under substitution. We start with sum and product. Recall the ranked alphabet  $\Delta$  defined in Section 4.

The next three Propositions show that algebraic linear orderings are closed under sum, product, and geometric sum. They were proved in [BE10b] for algebraic well-orderings. The same arguments work in the slightly more general setting of algebraic linear orderings. We repeat them for the reader's convenience. (The reader is invited to prove the same closure properties using deterministic context-free languages.) Assume that  $P$  and  $Q$  are respectively the leaf orderings of the trees  $T, S$  in  $T_\Delta^\omega$ .

**Proposition 6.1** *If  $P$  and  $Q$  are algebraic linear orderings, so is  $P + Q$ .*

*Proof.* Consider the tree  $+(a, b)$ , where  $a, b$  are letters of rank 0. Then the frontier of the tree  $+(T, S) = +(a, b)[a \mapsto T, b \mapsto S]$  is isomorphic to  $P + Q$ . Moreover, when  $T, S$  are algebraic, then so is  $+(T, S)$ .  $\square$

**Proposition 6.2** *If  $P$  and  $Q$  are algebraic linear orderings, so is  $Q \times P$ .*

*Proof.* The frontier of the tree  $T[1 \mapsto S]$  obtained by substituting a copy of  $S$  for each leaf of  $T$  isomorphic to  $Q \times P$ ; it is algebraic if  $T$  and  $S$  are.  $\square$

**Proposition 6.3** *If  $P$  is an algebraic linear ordering, then so is  $\sum_{n \geq 0} P^n$ .*

*Proof.* Let  $T'$  result from  $T$  by relabeling each leaf by the variable  $x_0$ . Consider the tree  $T_0$  of Example 4.6 and let  $S$  be the tree obtained by substituting the tree  $T'$  for each vertex labeled  $f$ :  $S = T_0[f \mapsto T']$ . Then the frontier of  $S$  is isomorphic to  $\sum_{n \geq 0} P^n$ . Moreover, if  $T$  is algebraic then so is  $S$ .  $\square$

Since for any (countable) ordinal  $\alpha > 1$ , the ordinal  $\alpha^\omega$  is  $\sum_{n \geq 0} \alpha^n$ , we immediately obtain:

**Corollary 6.4** *The collection of algebraic ordinals is closed under sum, product, and  $\omega$ -power.*

**Corollary 6.5** *Every ordinal less than  $\omega^\omega$  is algebraic.*

*Proof.* The ordinal 0 is algebraic, since if  $t$  is the empty tree in  $T_\Delta^\omega$ , then  $\text{Fr}(t)$  is the empty language whose frontier represents the ordinal 0. The ordinal 1 is algebraic since the tree **1** is algebraic. Thus the result follows from Proposition 2.11 and the closure properties of algebraic ordinals, Corollary 6.4.  $\square$

For the facts in Corollary 6.4 and Corollary 6.5, see also [BE07].

We establish two more closure properties of algebraic linear orderings. To prove these results, we will rely on the characterization of algebraic linear orderings as lexicographic orderings of dcl's.

**Proposition 6.6** *If  $P$  is an algebraic linear ordering and  $I$  is an interval of  $P$ , then  $I$  is algebraic.*

*Proof.* Since  $P$  is algebraic, it can be represented as the lexicographic ordering of a deterministic context-free language  $L \subseteq \{0, 1\}^*$ . If  $y_0 \in \{0, 1\}^*$ , then consider the languages

$$\begin{aligned} R_{<_\ell u_0} &= \{u \in \{0, 1\}^* : u <_\ell y_0\} \\ R_{\leq_\ell u_0} &= \{u \in \{0, 1\}^* : u \leq_\ell y_0\} \\ R_{>_\ell u_0} &= \{u \in \{0, 1\}^* : u >_\ell y_0\} \\ R_{\geq_\ell u_0} &= \{u \in \{0, 1\}^* : u \geq_\ell y_0\}. \end{aligned}$$

Clearly, all of them are regular. Since  $I$  can be represented as the lexicographic ordering of a finite intersection of  $L$  with such languages, and since the intersection of a deterministic context-free language with a regular language is deterministic context-free, it follows that  $I$  is algebraic.  $\square$

**Remark 6.7** *By Proposition 6.6, it follows that whenever  $\alpha$  is an algebraic ordinal, then so is any ordinal less than  $\alpha$ . Using this fact, we can give an alternative proof of Corollary 6.5. Indeed, as noted above, all ordinals of the form  $\omega^{\omega^n}$  are algebraic, for  $n \geq 0$ . Since each ordinal less than  $\omega^{\omega^\omega}$  is less than  $\omega^{\omega^n}$ , for some  $n$ , all ordinals less than  $\omega^{\omega^\omega}$  are algebraic.*

**Proposition 6.8** *The class of algebraic linear orderings is closed under the reverse operation.*

*Proof.* Suppose that  $P$  is an algebraic linear ordering that is isomorphic to the ordering  $(L, <_\ell)$  where  $L$  is a dcl over the ordered alphabet  $A = \{a_0 < \dots < a_{n-1}\}$ . Then let  $B = \{b_0, \dots, b_{n-1}\}$  be ordered by  $b_{n-1} < \dots < b_0$ , and let  $h : A^* \rightarrow B^*$  denote the homomorphism with  $h(a_i) = b_i$ , for all  $i \in [n]$ . Then  $h(L)$  is a dcl and  $P^*$  is isomorphic to  $(h(L), <_\ell)$ .  $\square$

## 7 Prefix and Scattered Grammars

In our proof of the fact that the Hausdorff rank of every scattered algebraic linear ordering is less than  $\omega^\omega$  we will make use of a corresponding result regarding the lexicographic ordering of languages generated by certain context-free grammars. For all unexplained (but standard) notions regarding context-free grammars and languages we refer to [HU79].

Below we will consider context-free grammars  $G = (N, \{0, 1\}, P, S)$  with set of nonterminals  $N$ , terminal alphabet  $\{0, 1\}$ , productions  $P$  and start symbol  $S$ . We will assume that either  $\mathcal{L}(G)$ , the language generated by  $G$ , is nonempty and  $G$  contains no useless nonterminals, or  $N = \{S\}$  and  $P$  is empty. For each  $p \in (N \cup \{0, 1\})^*$ , we let  $\mathcal{L}(p)$  denote the language

$$\mathcal{L}(p) = \{w \in \{0, 1\}^* : p \Rightarrow^* w\}$$

so that, in particular,  $\mathcal{L}(G) = \mathcal{L}(S)$ .

The nonterminals of a context-free grammar  $G = (N, \{0, 1\}, P, S)$  may be classified into strong components in the usual way. We recall the necessary definitions.

**Definition 7.1** *Suppose  $X, Y$  are nonterminals. Write*

$$Y \preceq X$$

*if there is a derivation  $X \Rightarrow^* pYq$  for some  $p$  and  $q$  in  $(N \cup \{0, 1\})^*$ . Define*

$$X \approx Y \iff X \preceq Y \text{ and } Y \preceq X.$$

The relation  $\preceq$  is a preorder on the nonterminals, and induces a partial order on the equivalence classes

$$[X] := \{Y : X \approx Y\},$$

where  $[Y] \leq [X]$  if  $Y \preceq X$ .

**Definition 7.2** The *height* of a nonterminal  $X$  is the number of equivalence classes  $[Y]$  strictly below  $[X]$  in this ordering.

We note some elementary properties of height.

**Proposition 7.3** Suppose  $X, Y$  are nonterminals.

- If  $X \approx Y$ , then  $X$  and  $Y$  have the same height.
- If the nonterminal  $X$  has height  $h$  and  $Y \preceq X$ , then  $Y$  has height at most  $h$ .
- If both  $X, Y$  have height  $h$  and if  $Y \preceq X$ , then  $X \approx Y$ .

□

In the proof of Theorem 7.15, we make use of the following easy lemma.

**Lemma 7.4** Suppose  $X \Rightarrow^* up$  is a (leftmost) derivation, where  $u \neq \epsilon \in \{0, 1\}^*$  and  $p \in (N \cup \{0, 1\})^*$ . If the nonterminal  $Y$  occurs in  $p$  and  $X \approx Y$ , then there is a (leftmost) derivation

$$X \Rightarrow^* uvXq,$$

for some  $v \in \{0, 1\}^*$  and  $q \in (N \cup \{0, 1\})^*$ .

□

Now we provide the definition of prefix grammars and scattered grammars that will play a crucial role.

**Definition 7.5** We call a context-free grammar  $G$  a **prefix grammar** if for each nonterminal  $X$ ,  $\mathcal{L}(X)$  is a prefix language. A **scattered grammar** is a prefix grammar  $G$  such that  $(\mathcal{L}(G), <_\ell)$  is a scattered linear ordering.

In the definition of scattered grammars, we only required that the lexicographic ordering of the language generated from the start symbol is scattered. As shown by the next result, it follows that the lexicographic ordering of the language generated from any nonterminal, and in fact any word possibly containing both nonterminals and terminals, is scattered.

**Proposition 7.6** *If  $G$  is a scattered grammar, then for each  $p \in (N \cup \{0, 1\})^*$ ,  $(\mathcal{L}(p), <_\ell)$  is a scattered linear ordering.*

*Proof.* We may assume  $\mathcal{L}(G) \neq \emptyset$ . For any  $p = u_0 X_1 u_1 \dots u_{n-1} X_n u_n$ , where each  $X_i$  is a nonterminal, and each  $u_i \in \{0, 1\}^*$ , the linear ordering  $(\mathcal{L}(p), <_\ell)$  is isomorphic to the cartesian product

$$\mathcal{L}(X_n, <_\ell) \times \dots \times \mathcal{L}(X_1, <_\ell),$$

(note the reverse order), since each language  $\mathcal{L}(X_i)$  is prefix. Thus, since the cartesian product of a finite number of scattered linear orderings is scattered, by Corollary 2.7, it suffices to prove that  $(\mathcal{L}(X), <_\ell)$  is scattered for each non-terminal  $X$ .

But for any nonterminal  $X$  there exist words  $u, v \in \{0, 1\}^*$  with  $S \Rightarrow^* uXv$ , and if  $u$  and  $v$  are such words, then, as above, the linear ordering  $(\mathcal{L}(X), <_\ell)$  is isomorphic to  $(u\mathcal{L}(X)v, <_\ell)$ , since  $\mathcal{L}(X)$  is a prefix language. Moreover, there is an order embedding of  $(u\mathcal{L}(X)v, <_\ell)$  into  $(\mathcal{L}(S), <_\ell)$ , which is a scattered linear ordering by assumption. Thus, by Lemma 2.3,  $(\mathcal{L}(X), <_\ell)$  is also scattered.  $\square$

The next lemma gives a condition on a context-free grammar  $G$  which implies that  $(\mathcal{L}(G), <_\ell)$  is not scattered.

**Lemma 7.7** *Suppose that  $G$  is a context-free grammar. Suppose that  $X \in N$  and  $u_i, v_i \in \{0, 1\}^*$  with  $X \Rightarrow^* u_i X v_i$ ,  $i = 0, 1, 2$  and  $u_0 <_s u_1 <_s u_2$ . Then  $(\mathcal{L}(X), <_\ell)$  is not scattered.*

*Proof.* Let  $w$  denote a word in  $\mathcal{L}(X)$ . Define

$$L_0 = \{u_{i_1} \dots u_{i_k} u_1 w v_1 v_{i_k} \dots v_{i_1} : i_1 \dots i_k \in \{0, 2\}^*, k \geq 0\}.$$

Clearly,  $L_0 \subseteq \mathcal{L}(X)$ . Also, if  $i_1 \dots i_k 1 <_s j_1 \dots j_m 1$ , then

$$u_{i_1} \dots u_{i_k} u_1 w v_1 v_{i_k} \dots v_{i_1} <_s u_{j_1} \dots u_{j_m} u_1 w v_1 v_{j_m} \dots v_{j_1}.$$

This shows that  $(\{0, 2\}^* 1, <_s)$  is isomorphic to  $(L_0, <_s)$ , which is in turn isomorphic to  $(L_0, <_\ell)$ . But it is easy to see that  $(\{0, 2\}^* 1, <_s)$  is isomorphic to the ordering of the rationals, so that  $(L_0, <_\ell)$  is not scattered.  $\square$

We use the above lemma together with Proposition 7.6 to prove:

**Proposition 7.8** *Suppose that  $G$  is a scattered grammar and  $X$  is a nonterminal. If  $X \Rightarrow^* uXp$  and  $X \Rightarrow^* vXq$  where  $u, v \in \{0, 1\}^*$  and  $p, q \in (N \cup \{0, 1\})^*$ , then either  $u \leq_p v$  or  $v \leq_p u$ .*

*Proof.* Let  $u', v' \in \{0, 1\}^*$  with  $p \Rightarrow^* u'$  and  $q \Rightarrow^* v'$ . We have the following derivations:

$$\begin{aligned} X &\Rightarrow^* uuXu'u' \\ X &\Rightarrow^* uvXv'u' \\ X &\Rightarrow^* vvXv'v' \end{aligned}$$

If neither  $u$  is a prefix of  $v$  nor  $v$  is a prefix of  $u$ , then  $u <_s v$  or  $v <_s u$ . In the first case,  $uu <_s uv <_s vv$ , while in the second  $vv <_s vu <_s uu$ , and thus by the previous lemma,  $(\mathcal{L}(X), <_\ell)$  is not scattered. This contradicts Proposition 7.6.  $\square$

*In the rest of this section, we will assume that  $G$  is a scattered grammar and for each nonterminal  $X$ ,  $\mathcal{L}(X)$  contains at least two words.*

It follows that  $G$  is  $\epsilon$ -free, i.e., there exists no production of the form  $X \rightarrow \epsilon$ . Since we may easily eliminate all chain productions  $X \rightarrow Y$ , where  $X, Y$  are nonterminals, we will also assume that  $G$  is free of chain productions.

**Proposition 7.9** *There exist no nonterminal  $X$  and words  $p, q, r \in (N \cup \{0, 1\})^*$  with  $X \Rightarrow^* pXqXr$ .*

*Proof.* Assume that  $X \Rightarrow^* pXqXr$ . Let  $u, v \in \{0, 1\}^*$  with  $p \Rightarrow^* u$  and  $q \Rightarrow^* v$ . Moreover, let  $x, y \in \mathcal{L}(X)$  be distinct words. We have that  $X \Rightarrow^* uxvXr$  and  $X \Rightarrow^* uyvXr$ . Now  $x <_s y$  or  $y <_s x$  and thus  $uxv <_s uyv$  or  $uyv <_s uxv$ , contradicting Proposition 7.8.  $\square$

A nonterminal  $X$  is *recursive* if there is a derivation  $X \Rightarrow^+ pXq$ , for some  $p, q \in (V \cup \{0, 1\})^*$ , and *left recursive* if there is a derivation  $X \Rightarrow^+ Xq$  for some  $q \in (V \cup \{0, 1\})^*$ .

**Proposition 7.10**  *$G$  is left-recursion free.*

*Proof.* Assume that  $X \Rightarrow^+ Xp$ . Since  $G$  is  $\epsilon$ -free and free of chain productions, we have that  $p \neq \epsilon$  and there exists some nonempty word  $u \in \mathcal{L}(p)$ . Let  $v \in \mathcal{L}(X)$ . Since  $vu$  is also in  $\mathcal{L}(X)$ , we conclude that  $\mathcal{L}(X)$  is not a prefix language, contradiction.  $\square$

Recall that a nonempty word  $u \in \{0, 1\}^*$  is *primitive* if it cannot be written in the form  $u = v^k$  for any  $v \in \{0, 1\}^*$  and integer  $k \geq 2$ . It is known, cf. [Lot97], Proposition 1.3.1, that each nonempty word  $u \in \{0, 1\}^*$  can be written in a unique way as  $u = v^k$ , where  $v$  is primitive, called the *primitive root* of  $u$ .

**Proposition 7.11** *For every recursive nonterminal  $X$  there is a primitive word  $u_0 \in \{0, 1\}^+$  such that whenever  $X \Rightarrow^+ uXp$  for some  $u \in \{0, 1\}^*$  and  $p \in (N \cup \{0, 1\})^*$ , then  $u = u_0^n$  for some  $n \geq 1$ .*



*Proof.* Suppose that  $X \Rightarrow^+ uXp$ , where  $u \in \{0,1\}^*$  and  $p \in (N \cup \{0,1\})^*$ . Since  $G$  is left recursion free, we have  $u \neq \epsilon$ . Let  $u_0$  be the primitive root of  $u$ . We will show that whenever  $X \Rightarrow^+ vXq$  with  $v \in \{0,1\}^+$  and  $q \in (N \cup \{0,1\})^*$  then there is some integer  $n$  with  $u_0^n = v$ , i.e.,  $u_0$  is also the primitive root of  $v$ .

So assume that  $X \Rightarrow^+ vXq$ . There exist integers  $k, m \geq 1$  with  $|u^k| = |v^m|$ . Since  $X \Rightarrow^+ u^k X p^k$  and  $X \Rightarrow^+ v^m X q^m$  and since  $|u^k| = |v^m|$ , we have that  $u^k = v^m$ , by Proposition 7.8. Thus  $v^m$  is a power of  $u_0$ , and since  $u_0$  is primitive,  $v$  also must be a power of  $u_0$ , see [Lot97].  $\square$

Sometimes we will write  $u_0^X$  for  $u_0$ .

**Definition 7.12** *Let  $X$  be a recursive nonterminal,  $u_0 = u_0^X$ . Then for each  $n \geq 0$  and prefix  $u1$  of  $u_0$  (so that  $u0$  is not a prefix of  $u_0$ ), let*

$$L(X, n, u1) = \{u_0^n u0v \in \{0,1\}^* : X \Rightarrow^* u_0^n u0v\}.$$

*Similarly, for each  $n \geq 0$  and prefix  $u0$  of  $u_0$ , let*

$$R(X, n, u0) = \{u_0^n u1v \in \{0,1\}^* : X \Rightarrow^* u_0^n u1v\}.$$

*Moreover, let*

$$L(X, n) = \bigcup_{u1 \leq_p u0} L(X, n, u1)$$

$$R(X, n) = \bigcup_{u0 \leq_p u0} R(X, n, u0).$$

**Proposition 7.13** • *If  $n < m$ , and  $x \in L(X, n)$ ,  $y \in L(X, m)$ , then  $x <_s y$ .*

- *If  $n < m$ , and  $x \in R(X, n)$ ,  $y \in R(X, m)$ , then  $y <_s x$ .*
- *For any  $n, m$  and words  $x \in L(X, n)$  and  $y \in R(X, m)$  it holds that  $x <_s y$ .*
- *Finally,  $\mathcal{L}(X) = L \cup R$ , where  $L = \bigcup_n L(X, n)$  and  $R = \bigcup_n R(X, n)$ .*

*Proof.* It is easy to check the first three claims.

Consider now a left derivation  $X \Rightarrow^* w$  of some word  $w \in \{0,1\}^*$ . Let  $k$  be large enough so that  $u_0^k$  is longer than  $w$ . Since  $\mathcal{L}(X)$  is a prefix language and contains a word of the form  $u_0^k x$ , it follows that  $w <_s u_0^k$  or  $u_0^k <_s w$ . In the first case, we can write  $w$  in the form  $w = u_0^n u0v$  for some words  $u, v$  such that  $u1$  is a prefix of  $u_0$ , so that  $w \in L(X, n, u1)$ . In the second case,  $w = u_0^n u1v$  for some words  $u, v$  such that  $u0$  is a prefix of  $u_0$  and thus  $w \in R(X, n, u0)$ .  $\square$

**Corollary 7.14**  $(\mathcal{L}(X), <_\ell)$  is isomorphic to the sum  $L + R$ , where  $L, R$  are the generalized sums

$$L = \sum_{n \geq 0} (L(X, n), <_\ell) \quad (4)$$

$$R = \sum_{n \leq 0} (R(X, -n), <_\ell). \quad (5)$$

*Proof.* Immediate, from the previous proposition.  $\square$

The following theorem is one of our main results.

**Theorem 7.15** Suppose that  $X$  is a nonterminal in  $G$  of height  $h$ . Then

$$r((\mathcal{L}(X), <_\ell)) \leq \omega^h + 1.$$

*Proof.* Suppose the height of  $X$  is 0. Note that if  $Y \preceq X$ , then  $X \approx Y$  by Proposition 7.3, so that if  $Y \neq X$  then  $X$  is recursive. Thus, if  $X$  is not recursive, since  $X$  has height 0,  $\mathcal{L}(X)$  is finite, and

$$r((\mathcal{L}(X), <_\ell)) \leq 1.$$

If  $X$  is recursive,  $\mathcal{L}(X)$  is isomorphic to  $L + R$  where  $L$  and  $R$  are defined in (4) and (5). We claim that for any  $n \geq 0$ , each language  $L(X, n)$  and  $R(X, n)$  is finite. Indeed, to prove  $L(X, n)$  is finite, it is enough to prove that if  $u1 \leq_p u_0$ , then  $L(X, n, u1)$  is finite, since  $L(X, n) = \bigcup_{u1 \leq_p u_0} L(X, n, u1)$ , where  $u$  ranges over all words in  $\{0, 1\}^*$  such that  $u1$  is a prefix of  $u_0 = u_0^X$ . Thus, suppose

$$X \Rightarrow^* u_0^n u_0 p$$

is a leftmost derivation, where  $p \in (N \cup \{0, 1\})^*$ . We claim that  $p$  cannot contain a nonterminal. Otherwise, if the nonterminal  $Y$  occurs in  $p$ ,  $X \approx Y$ , as noted above, so by Lemma 7.4, there is a derivation  $X \Rightarrow^* u_0^n u_0 w X q$ , contradicting Proposition 7.11. Thus,  $p$  is a word in  $\{0, 1\}^*$ . There is a finite number of leftmost such derivations with  $p \in \{0, 1\}^*$ , showing  $L(X, n, u1)$  is finite.

The argument showing  $R(X, n)$  is finite is the same. Thus,  $(\mathcal{L}(X), <_\ell)$  is isomorphic to  $\omega + \omega^*$ , so that  $r((\mathcal{L}(X), <_\ell)) \leq 2 = \omega^0 + 1$ .

Assume now that  $h > 0$ .

*Case 1.*  $X$  is not recursive. Then whenever  $X \rightarrow p$  is a production, any nonterminal occurring in  $p$  is of height less than  $h$ , by Proposition 7.3. By the induction hypothesis and Corollary 2.9, it follows that the Hausdorff rank of each  $(\mathcal{L}(p), <_\ell)$  such that  $X \rightarrow p$  is a production is less than  $\omega^h$ . Since  $\mathcal{L}(X)$  is a finite union of such languages  $\mathcal{L}(p)$ , by Corollary 2.15, the Hausdorff rank of  $(\mathcal{L}(X), <_\ell)$  is less than  $\omega^h$ .

*Case 2.*  $X$  is recursive. Then, by Corollary 7.14,  $(\mathcal{L}(X), <_\ell)$  is isomorphic to the sum  $L + R$ , where  $L$  and  $R$  are defined in (4) and (5). Below we will show that the rank of each  $(L(X, n), <_\ell)$  is less than  $\omega^h$ . The rank of each  $(R(X, -n), <_\ell)$  is also less than  $\omega^h$ , as may be shown in the same way, so that, by Corollary 2.9, the rank of  $(\mathcal{L}(X), <_\ell)$  is at most  $\omega^h + 1$ .

Now for each fixed  $n$ ,  $L(X, n)$  is itself a finite union  $\bigcup_{u1 \leq_p u_0} L(X, n, u1)$ , where  $u$  ranges over all words in  $\{0, 1\}^*$  such that  $u1$  is a prefix of  $u_0 = u_0^X$ . Since  $G$  is  $\epsilon$ -free and free of left recursion, for each fixed prefix  $u1$ ,  $L(X, n, u1)$  is a *finite* union of languages of the form  $\mathcal{L}(u_0^n u_0 p)$ , where there is a left derivation  $X \Rightarrow^+ u_0^n u_0 p$ . It follows that the nonterminals occurring in any such  $p$  have height less than  $h$ . Indeed, if a nonterminal  $Y$  with height  $h$  occurs in  $p$ , then  $X \approx Y$  and by Lemma 7.4, there is a derivation  $X \Rightarrow^+ u_0^n u_0 v X r$ , where  $v$  is a terminal word. But  $u_0^n u_0 v$  is not a power of  $u_0$ , contracting Proposition 7.11. Thus, by Corollary 2.9, we obtain that the rank of any  $(\mathcal{L}(u_0^n u_0 p), <_\ell)$  is less than  $\omega^h$ . Thus, by Corollary 2.15, it follows that  $r((L(X, n), <_\ell)) < \omega^h$ , for all  $n$ .  $\square$

**Corollary 7.16** *If  $G$  is a scattered grammar,  $r((\mathcal{L}(G), <_\ell)) < \omega^\omega$ .*

*Proof.* If the number of nonterminals in  $G$  is  $n$ , then the height of any nonterminal is less than  $n$ .  $\square$

## 8 From Algebraic Trees to Prefix Grammars

Consider a system of equations

$$F_i(x_0, \dots, x_{n_i-1}) = t_i(x_0, \dots, x_{n_i-1}), \quad i = 1, \dots, m. \quad (6)$$

where each  $t_i$  is a term over the ranked alphabet  $\Sigma \cup \mathcal{F}$  in the variables  $x_0, \dots, x_{n_i-1}$ . We assume that  $F_1$  is the principal function variable and that  $n_1 = 0$ . Each component of the least solution is an algebraic tree.

Let  $T$  be a finite or infinite tree in  $T_{\Sigma \cup \mathcal{F}}^\omega(V)$ . The labeled frontier languages of trees in  $T_{\Sigma \cup \mathcal{F}}^\omega(V)$  are defined as above.

(As mentioned above, it is known that a locally finite tree  $T$  is algebraic iff  $\text{Lfr}(T)$  is a deterministic context-free language.) We recall from [BE10b] the following result:

**Theorem 8.1** *When  $T \in T_\Sigma^\omega$  is an algebraic tree,  $\text{Lfr}(T)$  can be generated by a prefix grammar.*

Let  $(T = T_1, T_2, \dots, T_m)$  denote the least solution of the system (6). For the reader's convenience we recall from [BE10b] the construction of a prefix grammar generating  $\text{Lfr}(T)$ .

We will define a grammar whose nonterminals  $N$  consist of the letters  $F_i$ , together with all ordered pairs  $(F_i, j)$  where  $i = 1, \dots, m$ ,  $j \in [n_i]$ . The alphabet  $\Gamma$  of terminal letters is  $\Gamma = \Sigma_0 \cup \{(\gamma, j) : \gamma \in (\Sigma \cup \mathcal{F})_k, j \in [k]\}$ . The grammar is designed to have the following properties.

**Claim:** *For any word  $u$ ,  $T_i(u) = x_j$  iff  $(F_i, j) \Rightarrow^* \hat{u}$ . And for any word  $u$ ,  $T_i(u) \in \Sigma_0$  iff  $F_i \Rightarrow^* \hat{u}T_i(u)$ . Moreover, any terminal word derivable from  $(F_i, j)$  is of the form  $\hat{u}$ , and any terminal word derivable from  $F_i$  is of the form  $\hat{u}T_i(u)$  for some  $u \in \text{dom}(T_i)$ .*

The grammar generating  $\text{Lfr}(T)$  is:  $G_L = (N, \Gamma, P, F_1)$ , where the set  $P$  of productions is defined below. If  $t_1, \dots, t_m$  are the terms on the right side of (6) above, then the productions are:

- $$(F_i, j) \rightarrow \hat{u}$$

where  $u \in \text{dom}(t_i)$  and  $t_i(u) = x_j$ ,
- $$F_i \rightarrow \hat{u}t_i(u)$$

where  $u \in \text{dom}(t_i)$  and  $t_i(u) \in \Sigma_0 \cup \mathcal{F}$ .

The proof of the fact that this grammar is a prefix grammar generating the language  $\text{Lfr}(T)$  can be found in [BE10b].  $\square$

**Corollary 8.2** *If  $(Q, <)$  is an algebraic linear ordering, there is a prefix grammar  $G'$  with  $(Q, <)$  isomorphic to  $(\mathcal{L}(G'), <_\ell)$ .*

**Remark 8.3** *The above constructions can be carried out in polynomial time. Thus, each recursion scheme (over  $\Delta$ ) defining an algebraic linear ordering can be transformed in polynomial time into a prefix grammar defining the same linear ordering.*

## 9 Completing the Proof

To finish the argument, we apply Corollary 8.2.

**Theorem 9.1** *The Hausdorff rank of any scattered algebraic linear order is less than  $\omega^\omega$ .*

*Proof.* Indeed, any scattered algebraic linear ordering is isomorphic to the leaf ordering of an algebraic tree, by Proposition 4.4 above. By Corollary 8.2, there

is a prefix grammar  $G$  such that this leaf ordering is isomorphic to  $(\mathcal{L}(G), <_\ell)$ . It follows that  $G$  is a scattered grammar. By Corollary 7.16,  $r((\mathcal{L}(G), <_\ell)) < \omega^\omega$ .  $\square$

**Corollary 9.2 ([BE10b])** *The algebraic ordinals are precisely those less than  $\omega^{\omega^\omega}$ .*

*Proof.* Every ordinal less than  $\omega^{\omega^\omega}$  is algebraic, by Corollary 6.5. We need to show there are no more algebraic ordinals. But, if  $(Q, <)$  is an algebraic well-ordering of order type  $\alpha$ , then  $r(\alpha) < \omega^\omega$ , by Corollary 9.1. But then  $\alpha < \omega^{\omega^\omega}$ , by Corollary 2.4.  $\square$

## 10 Conclusion and Open Problems

A hierarchy of recursion schemes was introduced in [Ind76], see also [Damm77, Damm82, Gal84, Ong07, HMOS07], and many others. Here, we dealt with level 0 (regular schemes) and level 1 (algebraic or first-order schemes) of the hierarchy. In Theorem 9.1, we have shown that every scattered linear ordering definable by a level 1 scheme has Hausdorff rank less than  $\omega^\omega$ , whereas it has been known that the Hausdorff rank of any scattered linear ordering definable by a recursion scheme of level 0 is less than  $\omega$ . We conjecture that for each  $n$ , the Hausdorff rank of any scattered linear ordering definable by a level  $n$  scheme is less than  $\uparrow(\omega, n+1)$ , a tower of  $n+1$   $\omega$ 's. If that conjecture is true, then it follows that an ordinal is definable by a level  $n$  scheme if and only if it is less than  $\uparrow(\omega, n+2)$ , and thus an ordinal is definable in the hierarchy if and only if it is less than  $\epsilon_0$ . (See also [Braud], where it is shown that any ordinal less than  $\epsilon_0$  is definable in the hierarchy.)

In ordinal analysis of logical theories, the strength of a theory is measured by ordinals. For example, the proof theoretic ordinal of Peano arithmetic is  $\epsilon_0$ . Here we have a similar phenomenon: we measure the strength of recursive definitions by ordinals, and we conjecture that the ordinals definable are exactly those less than  $\epsilon_0$ . We also conjecture that the Hausdorff rank of any scattered linear ordering definable in the hierarchy of recursion schemes is less than  $\epsilon_0$ . The same may hold for the Caucal's pushdown hierarchy, [Cau03].

Finally, we mention two more open problems.

**Problem** Characterize the context-free well orderings and scattered linear orderings.

**Problem** Is it decidable for two algebraic linear orderings (each specified by a recursion scheme) whether they are isomorphic?

Put in other way, the question is whether it is decidable for two dcfl's equipped with the lexicographic ordering whether they are isomorphic.

## References

- [Ada74] J. Adamek. Free algebras and automata realizations in the language of categories. *Comment. Math. Univ. Carolinae*, 15(1974), 589–602.
- [AU72] A.V. Aho and J.D. Ullman, *The Theory of Parsing, Translation, and Compiling. Vol. I: Parsing*, Prentice-Hall, 1972.
- [BC01] S.L. Bloom and C.Choffrut. Long words: the theory of concatenation and  $\omega$ -power. *Theoretical Computer Science*, 259(2001), 533–548.
- [BE04] S.L. Bloom and Z. Ésik. Deciding whether the frontier of a regular tree is scattered. *Fundamenta Informaticae*, XI(2004) 1–22.
- [BE05] S.L. Bloom and Z. Ésik. The equational theory of regular words. *Information and Computation*, 197(2005), 55–89.
- [BE07] S.L. Bloom and Z. Ésik. Regular and algebraic words and ordinals. In: *CALCO 2007, Bergen*, LNCS 4624, Springer, 2007, 1–15.
- [BE10a] S.L. Bloom and Z. Ésik. A Mezei-Wright theorem for categorical algebras. *Theoretical Computer Science*, 411 (2010) 341–359.
- [BE10b] S.L. Bloom and Z. Ésik. Algebraic ordinals. To appear.
- [Braud] L. Braud. Unpublished paper, Laboratoire d’Informatique de l’Institut Gaspard Monge de l’Université Paris-Est à Marne-la-Vallée, 2008.
- [BKN08] M. Brough, B. Khoussainov, and P. Nelson. Sequential automatic algebras. In: A. Beckmann, C. Dimitracopoulos, and B. Löwe (Eds.): *CiE 2008*, LNCS 5028, pp. 84–93, 2008.
- [Cau03] D. Caucal, On infinite transition graphs having a decidable monadic theory, *Theoret. Comput. Sci.*, 290(2003), 79–115.
- [Col04] Th. Colcombet. Equational presentations of tree automatic structures. *Workshop on Automata, Structures, and Logic*, Auckland, 2004.
- [Cour78a] B. Courcelle. Frontiers of infinite trees. *RAIRO Theoretical Informatics and Applications*, 12(1978), 319–337.
- [Cour78] B. Courcelle. A representation of trees by languages, *Theoretical Computer Science*, 6 (1978), 255–279 and 7(1978), 25–55.
- [Cour83] B. Courcelle. Fundamental properties of infinite trees. *Theoretical Computer Science*, 25(1983), 95–169.
- [Damm77] W. Damm. Higher type program schemes and their tree languages. In: *Theoretical Computer Science (Third GI Conf., Darmstadt, 1977)*, LNCS 48, Springer, Berlin, 1977, 51–72.

- [Damm82] W. Damm. The IO and OI hierarchies. *Theoretical Computer Science*, 20(1982), 95–206.
- [DT90] M. Dauchet and S. Tison. The theory of ground rewrite systems is decidable. In *LICS 1990*, pages 242–248. IEEE, 1990.
- [Del04] Ch. Delhommé. Automaticity of ordinals and of homogeneous graphs. *C. R. Math. Acad. Sci. Paris* 339(2004), no. 1, 5–10. (in French)
- [Gal84] J. Gallier.  $n$ -rational algebras. I. Basic properties and free algebras. *SIAM J. Comput.*, 13(1984), 750–775.
- [GTWW77] J. A. Goguen, J. W. Thatcher, E. G. Wagner and J. B. Wright. Initial algebraic semantics and continuous algebras. *J. ACM*, 24(1977), 68–95.
- [GS84] F. Gécseg and M. Steinby. *Tree Automata* Akadémiai Kiadó, Budapest, 1984.
- [Gue81] I. Guessarian. *Algebraic Semantics*, LNCS 99, 1981.
- [Heil80] S. Heilbrunner. An algorithm for the solution of fixed-point equations for infinite words. *RAIRO Theoretical Informatics and Applications*, 14(1980), 131–141.
- [Hod82] B.R. Hodgson. On direct products of automaton decidable theories. *Theoretical Computer Science*, 19:331–335, 1982.
- [HMOS07] M. Hague, A.S. Murawawski, C.-H Luke Ong, O. Serre. Collapsible pushdown automata and recursion schemes. *23rd Annual IEEE Symposium on Logic in Computer Science, LICS*, IEEE, 2008, 452–461.
- [HU79] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley, 1979.
- [Ind76] K. Indermark, Schemes with recursion on higher types, in: *Proc. Mathematical Foundations of Computer Science*, LNCS 45(1976), 325–358.
- [Jonst87] P.T. Johnstone. *Notes on logic and set theory* Cambridge Mathematical Textbooks 1987.
- [KN95] B. Khoussainov, A. Nerode. Automatic presentations of structures. in: *Logic and Comput. Complex., Lecture Notes in Comput. Sci.*, vol. 960, 1995, pp. 367–392.
- [KRS03] B. Khoussainov, S. Rubin and F. Stephan. On automatic partial orders. *Proceedings of Eighteenth IEEE Symposium on Logic in Computer Science, LICS*, 168–177, 2003.
- [Lot97] M. Lothaire. *Combinatorics on Words*, Cambridge University Press 1983, 1997.

- [MW67] J. Mezei and J. Wright. Algebraic automata and context-free sets. *Information and Control*, 11 (1967), 3–29.
- [Ong07] C.-H Luke Ong. Hierarchies of infinite structures generated by pushdown automata and recursion schemes. *MFCS 2007*, LNCS 4708, Springer, 2007, 15–21.
- [Roit90] J. Roitman. *Introduction to Modern Set Theory*. Wiley, 1990.
- [Ros82] J.B. Rosenstein. *Linear Orderings*. Academic Press, New York, 1982.
- [Thom86] W. Thomas. On frontiers of regular trees. *RAIRO Theoretical Informatics and Applications*, 20(1986), 371–381.
- [Wand79] M. Wand. Fixed point constructions in order-enriched categories. *Theoretical Computer Science*, 8(1979), 13–30.