

# First-Order logic and its **Infinitary Quantifier Extensions over Countable Words**

Bharat Adsul\*, Saptarshi Sarkar\*, and A.V. Sreejith<sup>†</sup>

\*IIT Bombay, India and <sup>†</sup>IIT Goa, India

**Abstract.** We contribute to the refined understanding of the language-logic-algebra interplay in the context of *first-order* properties of countable words. We establish decidable algebraic characterizations of one variable fragment of FO as well as boolean closure of existential fragment of FO via a strengthening of Simon’s theorem about piecewise testable languages. We propose a new extension of FO which admits infinitary quantifiers to reason about the inherent infinitary properties of countable words. We provide a very natural and hierarchical block-product based characterization of the new extension. We also explicate its role in view of other natural and classical logical systems such as WMSO and FO[cut] - an extension of FO where quantification over Dedekind-cuts is allowed. We also rule out the possibility of a finite-basis for a block-product based characterization of these logical systems. Finally, we report simple but novel algebraic characterizations of one variable fragments of the hierarchies of the new proposed extension of FO.

**Keywords:** Countable words · First-order logic · Monoids.

## 1 Introduction

Over finite words, we have a foundational language-logic-algebra connection (see [18,10]) which equates regular-expressions, MSO-logic, and (recognition by) finite monoids/automata. In fact, one can effectively associate, to a regular language, its finite *syntactic monoid*. This canonical algebraic structure carries a rich amount of information about the corresponding language. Its role is highlighted by the classical Schutzenberger-McNaughton-Papert theorem (see, for instance, [11]) which shows that *aperiodicity* property of the syntactic monoid coincides with describability using star-free expressions as well as definability in First-Order (FO) logic. So, we arrive at a refined understanding of the language-logic-algebra connection to an important subclass of regular languages: it equates star-free regular expressions, FO-logic, and aperiodic finite monoids.

A variety of algebraic tools have been developed and crucially used to obtain deeper insights. Some of these tools [11,15,17] are: ordered monoids, the so-called Green’s relations, wreath/block products and related principles etc. Let us mention Simon’s celebrated theorem [14] - which equates piecewise-testable languages, Boolean closure of the existential fragment of FO-logic and  $J$ -trivial finite monoids<sup>1</sup>. It is important to note that this is an effective characterization,

---

<sup>1</sup> It refers to  $J$  - one of the fundamental Green’s relations

that is, they provide a decidable characterization of the logical fragment. There have been several results of this kind (see the survey [6]). Another particularly interesting set of results is in the spirit of the fundamental Krohn-Rhodes theorem. These results establish a block-product based decompositional characterization of a logical fragment and have many important applications [15]. The prominent examples are a characterization of FO-logic (resp.  $\text{FO}^2$ , the two-variable fragment) in terms of strongly (resp. weakly) iterated block-products of copies of the unique 2-element aperiodic monoid.

One of the motivations for this work is to establish similar results in the theory of regular languages of **countable words**. We use the overarching algebraic framework developed in the seminal work [4] to reason about languages of **countable words**. This framework extends the language-logic-algebra interplay to the setting of countable words. It develops fundamental algebraic structures such as *finite*  $\otimes$ -**monoids** and  $\otimes$ -**algebras** and equates MSO-definability with **recognizability** by these algebraic structures. A detailed study of a variety of sub-logics of MSO over **countable words** is carried out in [5]. This study also extends classical Green’s relations to  $\otimes$ -**algebras** and makes heavy use of it. Of particular interest to us are the results about algebraic **equational** characterizations of FO, FO[cut] – an *extension* of FO that allows quantification over **Dedekind cuts** and WMSO – an *extension* of FO that allows quantification over finite sets. A decidable algebraic characterization of  $\text{FO}^2$  over **countable words** is also presented in [9]. Another recent development [1] is the seamless integration of **block products** into the countable setting. The work introduces the **block product** operation of the relevant algebraic structures and establishes an appealing **block product principle**. Further, it naturally extends the above-mentioned **block product** characterizations of FO and  $\text{FO}^2$  to **countable words**.

In this work, we begin our explorations into the *small* fragments of FO over **countable words**, guided by the choice of results in [6]. We arrive at the language-logic-algebra connection for  $\text{FO}^1$  – the one variable fragment of FO. Coupled with earlier results about  $\text{FO}^2$  and  $\text{FO}=\text{FO}^3$  (see [7]), this completes our algebraic understanding of FO fragments defined by the number of permissible variables. We next extend **Simon’s theorem** on **piecewise testable languages** to **countable words** and provide a natural algebraic characterization of the Boolean closure of the existential-fragment of FO. Fortunately or unfortunately, depending on the point of view, this landscape of small fragments of FO over **countable words** parallels very closely the same landscape over finite words. This can be attributed to the limited expressive power of FO over **countable words**. For instance, Bès and Carton [3] showed that the seemingly natural ‘finiteness’ property (that the set of all positions is a finite set) of **countable words** can not be expressed in FO!

One of the main contributions of this work is the introduction of new *infinitary* quantifiers to FO. The works [2,8] also extend FO over arbitrary structures by *cardinality/infinitary-counting* quantifiers and studies decidable theories thereof. An extension of FO over finite and  $\omega$ -words by *modulus-counting quantifiers* is algebraically characterized in [16]. The main purpose of our new quantifiers is to naturally allow expression of infinitary features which are inherent in

the countable setting and study the resulting definable formal languages in the algebraic framework of [4]. An example formula using such an infinitary quantifier is:  $\exists^{\infty_1} x : a(x) \wedge \neg \exists^{\infty_1} x : b(x)$ . In its natural semantics, this formula with one variable asserts that there are infinitely many  $a$ -labelled positions and only finitely many  $b$ -labelled positions. We propose an extension of FO called FO[ $\infty$ ] that supports *first-order infinitary quantifiers* of the form  $\exists^{\infty_k} x$  to talk about existence of higher-level infinitely (more accurately, *Infinitary rank  $k$* ) many witnesses  $x$ . We organize FO[ $\infty$ ] in a natural hierarchy based on the maximum allowed infinitary-level of the quantifiers.

We now summarize the key technical results of this paper. We establish a hierarchical *block product* based characterization of FO[ $\infty$ ]. Towards this, we identify an appropriate simple family of  *$\otimes$ -algebras* and show that this family (in fact, its initial fragments) serve as a basis in our hierarchical *block product* based characterization. We establish that FO[ $\infty$ ] properties can be expressed simultaneously in FO[cut] as well as WMSO. We also show that the language-logic-algebra connection for FO<sup>1</sup> admits novel generalizations to the one variable fragments of the new extension of FO. We finally present ‘no finite *block product* basis’ theorems for our FO extensions, FO[cut], and the class FO[cut]  $\cap$  WMSO. This is in contrast with [1] where the unique 2-element  *$\otimes$ -algebra* is a basis for a block-product based characterization of FO.

The rest of the paper is organized as follows. Section 2 recalls basic notions about *countable words* and summarizes the necessary algebraic background from the framework [4]. Section 3 deals with the small fragments of FO: FO<sup>1</sup> and the Boolean closure of the existential fragment of FO. Section 4 contains the extensions FO[ $\infty$ ] and results relevant to it. Section 5 is concerned with ‘no finite *block product* basis’ theorems.

## 2 Preliminaries

In this section we briefly recall the algebraic framework developed in [4].

**Countable words** A *countable linear ordering* (or simply ordering)  $\alpha = (X, <)$  is a non-empty countable set  $X$  equipped with a total order:  $X$  is the *domain* of  $\alpha$ . An ordering  $\beta = (Y, <)$  is called a *subordering* of  $\alpha$  if  $Y \subseteq X$  and the order on  $Y$  is induced from that of  $X$ . We denote by  $\omega, \omega^*, \delta, \eta$  the orderings  $(\mathbb{N}, <), (-\mathbb{N}, <), (\mathbb{Z}, <), (\mathbb{Q}, <)$  respectively. A *Dedekind cut* (or simply a *cut*) is a left-closed subset  $Y \subseteq X$  of  $\alpha$ . Given disjoint *linear orderings*  $(\beta_i)_{i \in \alpha}$  indexed with a *linear ordering*  $\alpha$ , their *generalized sum*  $\sum_{i \in \alpha} \beta_i$  is the *linear ordering* over the union of the domains of the  $\beta_i$ ’s, with the order defined by  $x < y$  if either  $x \in \beta_i$  and  $y \in \beta_j$  with  $i < j$ , or  $x, y \in \beta_i$  for some  $i$ , and  $x < y$  in  $\beta_i$ . The book [12] contains a detailed study of *linear orderings*.

An *alphabet*  $\Sigma$  is a finite set of symbols called *letters*. Given a *linear ordering*  $\alpha$ , a *countable word* (henceforth called *word*) over  $\Sigma$  of *domain*  $\alpha$  is a mapping  $w : \alpha \rightarrow \Sigma$ . The *domain* of a *word* is denoted  $dom(w)$ . For a subset  $I \subseteq dom(w)$ ,  $w|_I$  denotes the *subword* got by restricting  $w$  to the domain  $I$ . If  $I$  is an *interval*

( $\forall x, y \in I, x < z < y \rightarrow z \in I$ ) then  $w|_I$  is called a *factor* of  $w$ . The set of all words is denoted  $\Sigma^\otimes$  and the set of all non-empty (resp. finite) words  $\Sigma^\oplus$  (resp.  $\Sigma^*$ ). A *language* (of countable words) is a subset of  $\Sigma^\otimes$ . The *generalized concatenation* of the words  $(w_i)_{i \in \alpha}$  indexed by a *linear ordering*  $\alpha$  is  $\prod_{i \in \alpha} w_i$  and denotes the word  $w$  of domain  $\sum_{i \in \alpha} \beta_i$  where  $\beta_i$  are disjoint and such that  $w|_{\beta_i}$  is isomorphic to  $w_i$  for all  $i \in \alpha$ .

The *empty word*  $\varepsilon$ , is the only word of empty domain. The *omega power* of a word  $w$  is defined as  $w^\omega := \prod_{i \in \omega} w$ . The *omega\* power* of a word  $w$ , denoted by  $w^{\omega^*}$ , is the concatenation of omega\* many  $w$ 's. The *perfect shuffle* for a non-empty finite set of letters  $A \subseteq \Sigma$  (denoted by  $A^\eta$ ) is a word of domain  $(\mathbb{Q}, <)$  in which only letters from  $A$  occur and, all non-empty and non-singleton intervals contain at least one occurrence of each letter in  $A$ . This word is unique up to isomorphism [13]. We can extend the notion of perfect shuffle to a finite set of words  $W = \{w_1, \dots, w_k\}$ . We define  $W^\eta$  to be  $\prod_{i \in \mathbb{Q}} w_{f(i)}$  where  $f : (\mathbb{Q}, <) \rightarrow \{1, 2, \dots, k\}$  is the unique perfect shuffle over the set of letters  $\{1, 2, \dots, k\}$ .

**The algebra** A  $\otimes$ -monoid  $\mathbf{M} = (M, \mathbf{id}, \pi)$  is a set  $M$  equipped with an operation  $\pi$ , called the *product*, from  $M^\otimes$  to  $M$ , that satisfies  $\pi(a) = a$  for all  $a \in M$ , and the *generalized associativity* property: for every words  $u_i$  over  $M$  with  $i$  ranging over a countable linear ordering  $\alpha$ ,  $\pi(\prod_{i \in \alpha} u_i) = \pi(\prod_{i \in \alpha} \pi(u_i))$ . We reserve the notation  $\mathbf{id}$  for the *identity element*  $\mathbf{id} = \pi(\varepsilon)$ ; it is called the neutral element in [4]. An example of a  $\otimes$ -monoid is the free  $\otimes$ -monoid  $(\Sigma^\otimes, \varepsilon, \prod)$  over the alphabet  $\Sigma$  with the product being the generalized concatenation. Now we discuss some natural algebraic notions. A *morphism* from a  $\otimes$ -monoid  $(M, \mathbf{id}, \pi)$  to a  $\otimes$ -monoid  $(M', \mathbf{id}', \pi')$  is a map  $h : M \rightarrow M'$  such that, for every  $w \in M^\otimes$ ,  $h(\pi(w)) = \pi'(\bar{h}(w))$  where  $\bar{h}$  is the pointwise extension of  $h$  to words. We skip the notions sub- $\otimes$ -monoid and direct products since they are as expected. We say  $\mathbf{M} = (M, \mathbf{id}, \pi)$  *divides*  $\mathbf{M}' = (M', \mathbf{id}', \pi')$  if there exists a sub- $\otimes$ -monoid  $\mathbf{M}'' = (M'', \mathbf{id}'', \pi'')$  of  $\mathbf{M}'$  and a surjective morphism from  $\mathbf{M}''$  to  $\mathbf{M}$ .

A  $\otimes$ -monoid  $\mathbf{M} = (M, \mathbf{id}, \pi)$  is said to be finite if  $M$  is so. Note that, even for a finite  $\otimes$ -monoid, the product operation  $\pi$  has an infinitary description. It turns out that  $\pi$  can be captured using finitely presentable *derived operations*. Corresponding to a  $\otimes$ -monoid  $(M, \mathbf{id}, \pi)$  there is an induced  $\otimes$ -algebra  $\mathbf{M} = (M, \mathbf{id}, \cdot, \tau, \tau^*, \kappa)$  where the operations are defined as following: for all  $a, b \in M$ ,  $a \cdot b = \pi(ab)$ ,  $a^\tau = \pi(a^\omega)$ ,  $a^{\tau^*} = \pi(a^{\omega^*})$  and for all  $\emptyset \neq E \subseteq M$ ,  $E^\kappa = \pi(E^\eta)$ . These *derived operators* satisfy certain natural axioms; see [4] for more details. It has been established in [4] that an *arbitrary* finite  $\otimes$ -algebra  $\mathbf{M} = (M, \mathbf{id}, \cdot, \tau, \tau^*, \kappa)$  satisfying these natural axioms is induced by a unique  $\otimes$ -monoid  $\mathbf{M} = (M, \mathbf{id}, \pi)$ . We later introduce the notion of an *evaluation tree* which aids this correspondence. It is rather straightforward to define the notions of morphisms, subalgebras, direct-products as well as division for  $\otimes$ -algebras.

It follows from the definition of a  $\otimes$ -algebra  $\mathbf{M} = (M, \mathbf{id}, \cdot, \tau, \tau^*, \kappa)$  that  $(M, \mathbf{id}, \cdot)$  is a *monoid*, that is the operation  $\cdot$  is associative with identity  $\mathbf{id}$ . For a singleton set  $E = \{m\}$ , we write  $m^\kappa = \{m\}^\kappa$ . Notice that for all  $m \in M$ ,  $m \cdot \mathbf{id} = \mathbf{id} \cdot m = m$  and for all  $\emptyset \neq E \subseteq M$ ,  $E^\kappa = (E \cup \{\mathbf{id}\})^\kappa$ . Further,

$\mathbf{id}^\tau = \mathbf{id}^{\tau^*} = \mathbf{id}^\kappa = \mathbf{id}$ . As a result, in our definitions of  $\otimes$ -algebras later in the paper, we restrict the descriptions of **derived operators** to  $M \setminus \{\mathbf{id}\}$ .

An *evaluation tree* over a word  $u \in M^{\otimes} \setminus \{\varepsilon\}$  is a tree  $\mathcal{T} = (T, h)$  such that every branch/path of  $\mathcal{T}$  is of finite length and where every vertex in  $T$  is a factor of  $u$ , the root is  $u$  and  $h : T \rightarrow M$  is a map such that:

- A leaf is a singleton letter  $a \in M$  such that  $h(a) = a$ .
- Internal nodes have either two or  $\omega$  or  $\omega^*$  or  $\mathbb{Q}$  many children.
- If  $w$  has children  $v_1$  and  $v_2$ , then  $w = v_1 v_2$  and  $h(w) = h(v_1) \cdot h(v_2)$ .
- If  $w$  has  $\omega$  many children  $\langle v_1, v_2, \dots \rangle$ , then there is an *idempotent*<sup>2</sup>  $e$  such that  $e = h(v_i)$  for all  $i \geq 1$ , and  $w = \prod_{i \in \omega} v_i$  and  $h(w) = e^\tau$ .
- If  $w$  has  $\omega^*$  many children  $\langle \dots, v_{-2}, v_{-1} \rangle$ , then there is an *idempotent*  $f$  such that  $f = h(v_i)$  for all  $i \leq -1$ , and  $w = \prod_{i \in \omega^*} v_i$  and  $h(w) = f^{\tau^*}$ .
- If  $w$  has  $\mathbb{Q}$  many children  $\langle v_i \rangle_{i \in \mathbb{Q}}$ , then  $w = \prod_{i \in \mathbb{Q}} v_i$  where for the *perfect shuffle*  $f$  over an  $E = \{a_1, \dots, a_k\} \subseteq M$ ,  $h(v_i) = a_{f(i)}$ , and  $h(w) = E^\kappa$ .

The *value* of  $\mathcal{T}$  is defined to be  $h(u)$ . It was shown in [4, Proposition 8 and 9] that every word  $u$  has an *evaluation tree* and the *values* of two *evaluation trees* of  $u$  are equal and they are equal to  $\pi(u)$ . Therefore, a  $\otimes$ -algebra defines the *generalized associativity* product  $\pi : M^{\otimes} \rightarrow M$ . The correspondence between finite  $\otimes$ -monoids and  $\otimes$ -algebras permits interchangeability; we exploit it implicitly.

A morphism from the *free  $\otimes$ -monoid*  $\Sigma^{\otimes}$  to  $\mathbf{M}$  is described (determined) by a map  $h' : \Sigma \rightarrow M$ ; we simply write  $h' : \Sigma \rightarrow \mathbf{M}$ . With  $h'$  also denoting its pointwise extension  $h' : \Sigma^{\otimes} \rightarrow M^{\otimes}$ , given a word  $u \in \Sigma^{\otimes}$ , we can use the evaluation tree over the word  $h'(u) \in M^{\otimes}$  to obtain  $\pi(h'(u)) \in M$ . By further abuse of notation,  $h' : \Sigma^{\otimes} \rightarrow M$  also denotes the morphism which sends  $u$  to  $\pi(h'(u))$ . We say that  $L$  is recognized by  $\mathbf{M}$  if there exists a map/morphism  $h' : \Sigma^{\otimes} \rightarrow \mathbf{M}^{\otimes}$  such that  $L = h'^{-1}(h'(L))$ . The fundamental result of [4] states that *regular languages* (MSO definable languages) are exactly those *recognized* by finite  $\otimes$ -monoids (equivalently  $\otimes$ -algebras). It is important to note that, every regular language  $L$  is associated a finite (canonical/minimal) syntactic  $\otimes$ -monoid which divides every  $\otimes$ -monoid that recognizes  $L$ . Further, it can be represented as a  $\otimes$ -algebra from a finite description of  $L$ .

*Example 1.* The  $\otimes$ -monoid  $\mathbf{U}_1 = (\{\mathbf{id}, 0\}, \mathbf{id}, \pi)$  and its induced  $\otimes$ -algebra are shown on the left and right respectively.

$$\pi(u) = \begin{cases} \mathbf{id} & \text{if } u \in \{\mathbf{id}\}^{\otimes} \\ 0 & \text{otherwise} \end{cases} \quad \begin{array}{c|c|c|c} \mathbf{id} & 0 & \tau & \tau^* \\ \mathbf{id} & \mathbf{id} & 0 & \mathbf{id} & \mathbf{id} \\ 0 & 0 & 0 & 0 \end{array} \quad S^\kappa = \begin{cases} \mathbf{id} & \text{if } S = \{\mathbf{id}\} \\ 0 & \text{otherwise} \end{cases}$$

Let  $\Sigma = \{a, b\}$  and  $L$  be the set of words which contain an occurrence of letter  $a$ . It is easy to see that the map  $h : \Sigma \rightarrow \mathbf{U}_1$  sending  $h(a) = 0, h(b) = \mathbf{id}$  recognizes  $L$  as  $L = h^{-1}(0)$ . In fact,  $\mathbf{U}_1$  is the syntactic  $\otimes$ -monoid of  $L$ .

<sup>2</sup> An *idempotent* is an element  $e$  where  $e \cdot e = e$

*Example 2.* Consider the  $\otimes$ -algebra  $\text{Gap} = (\{\text{id}, [], (), g\}, \text{id}, \cdot, \tau, \tau^*, \kappa)$ . We let  $\Sigma = \{a\}$  and define the map  $h : \Sigma \rightarrow \text{Gap}$  as  $h(a) = []$ . The resulting morphism maps a word  $u$  to  $h(u) = g$  iff the word  $u$  admits a **gap**; that is a **cut** with no maximum and its complement has no minimum. Other words are mapped to their right ‘ends-type’: for instance,  $h(u) = []$  iff  $\text{dom}(u)$  has a minimum and no maximum. For a word  $v = a^\omega a^{\omega^*}$ , the pointwise extension  $v' = h(v) = []^\omega []^{\omega^*}$ . An example **evaluation tree**  $\mathcal{T}$  for  $v'$  consists of root with two children. The left (resp. right) child has  $\omega$  (resp.  $\omega^*$ ) many children  $[]$  and has **value**  $[]^\tau$  (resp.  $[]^{\tau^*}$ ). As a result, the value of  $\mathcal{T}$  is  $[]^\tau \cdot []^{\tau^*} = [] \cdot () = g$ .

$\cdot$	$[]$	$[]$	$()$	$()$	$g$	$\tau$	$\tau^*$
$[]$	$[]$	$[]$	$[]$	$[]$	$g$	$[]$	$()$
$[]$	$[]$	$[]$	$g$	$g$	$g$	$[]$	$()$
$()$	$()$	$()$	$()$	$()$	$g$	$()$	$[]$
$()$	$()$	$()$	$g$	$g$	$g$	$g$	$g$
$g$	$g$	$g$	$g$	$g$	$g$	$g$	$g$

$$S^\kappa = \begin{cases} \text{id} & \text{if } S = \{\text{id}\} \\ g & \text{otherwise} \end{cases}$$

We can characterize sets of  $\otimes$ -monoids using *identities*. For example, we say that  $\mathbf{U}_1$  satisfies *commutative* equation  $x \cdot y = y \cdot x$ . This means that the equation holds for any assignment of elements in the  $\otimes$ -monoids to the variables  $x$  and  $y$ . Like in the case of **monoids**, the set of  $\otimes$ -monoids satisfying a set of equations are closed under subsemigroup, **division** and *direct product* [5].

The *block product* of  $\otimes$ -monoids  $\mathbf{M}$  and  $\mathbf{N}$ , is denoted by  $\mathbf{M} \square \mathbf{N}$  and is the semidirect product of  $\mathbf{M}$  and  $\mathbf{K} = \mathbf{N}^{M \times M}$  with respect to the *canonical* left and right ‘action’ of  $\mathbf{M}$  on  $\mathbf{K}$ . The details are given in [1]. The *block product principle* characterizes languages defined by **block product** of  $\otimes$ -monoids. Towards this, fix a map  $h : \Sigma \rightarrow \mathbf{M} \square \mathbf{N}$  such that  $h(a) = (m_a, f_a)$  where  $m_a \in M$  and  $f_a : M \times M \rightarrow N$ . The map  $h_1 : \Sigma \rightarrow \mathbf{M}$  setting  $h_1(a) = m_a$  defines a morphism  $h_1 : \Sigma^* \rightarrow M$ . We define the *transducer*  $\sigma : \Sigma^* \rightarrow (M \times \Sigma \times M)^*$  as follows: let  $u \in \Sigma^*$  with domain  $\alpha$ . The word  $u' = \sigma(u)$  has domain  $\alpha$  and for a position  $x \in \alpha$ ,  $u'(x) = (h_1(u_{<x}), u(x), h_1(u_{>x}))$ . Here  $u_{<x}$  (resp.  $u_{>x}$ ) is the subword of  $u$  on positions strictly less (resp. greater) than  $x$ .

**Proposition 1 (Block Product Principle [1]).** *Let  $L \subseteq \Sigma^*$  be recognized by  $h : \Sigma \rightarrow \mathbf{M} \square \mathbf{N}$ . Then  $L$  is a boolean combination of languages of the form  $L_1$  and  $\sigma^{-1}(L_2)$  where  $L_1$  and  $L_2$  are recognized by  $\mathbf{M}$  and  $\mathbf{N}$  respectively and  $\sigma : \Sigma^* \rightarrow (M \times \Sigma \times M)^*$  is a state-based transducer.*

### 3 Small fragments of FO

In this section, we focus on two particularly small fragments of first-order logic interpreted over countable words. First-order logic uses variables  $x, y, z, \dots$  which are interpreted as positions in the domain of a word. The syntax of *first-order logic* (FO) is:  $x < y \mid a(x) \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid \neg \phi \mid \exists x \phi$ , for all  $a \in \Sigma$ .

We skip the natural semantics. A **language**  $L$  of **countable words** is said to be **FO-definable** if there exists an **FO-sentence**  $\phi$  such  $L = \{u \in \Sigma^* \mid u \models \phi\}$ .

Recall that the classical Schutzenberger-McNaughton-Papert theorem characterizes FO-definability of a regular language of finite words in terms of aperiodicity of its finite syntactic monoid. The survey [6] presents similar decidable characterizations of several interesting small fragments of FO-logic such as  $\text{FO}^1$ ,  $\text{FO}^2$ ,  $B(\exists^*)$  – boolean closure of the existential first-order logic. It is known [7] that, over finite *as well as countable words*,  $\text{FO} = \text{FO}^3$ . As mentioned in the introduction, over countable words, we already have decidable algebraic characterizations of  $\text{FO}^3$  from [5] and  $\text{FO}^2$  from [9]. Here we identify decidable algebraic characterizations, over *countable words*, for  $\text{FO}^1$  and  $B(\exists^*)$ .

### 3.1 FO with single variable

The fragment  $\text{FO}^1$  has access to only one variable. We recall that over finite words a *regular language* is  $\text{FO}^1$ -definable iff its syntactic monoid is *commutative* and *idempotent*. We henceforth focus our attention to  $\text{FO}^1$  on *countable words*.

Clearly,  $\text{FO}^1$  can recognize all words with a particular *letter*. With a single variable the logic cannot talk about order of *letters* or count the number of occurrence of a *letter*. This gives an intuition that the syntactic  $\otimes$ -monoid of a language definable in  $\text{FO}^1$  is *commutative* and *idempotent*.

We say that a  $\otimes$ -algebra  $\mathbf{M} = (M, \text{id}, \cdot, \tau, \tau^*, \kappa)$  is *shuffle-trivial* if it satisfies the equational *identity*:

$$\{x_1, \dots, x_p\}^\kappa = x_1 \cdot x_2 \cdot \dots \cdot x_p$$

Then  $\mathbf{M}$  is commutative:  $x \cdot y = \{x, y\}^\kappa = \{y, x\}^\kappa = y \cdot x$ . Moreover, every element of  $\mathbf{M}$  is a *shuffle-idempotent*: for all  $m \in M$ ,  $m^\kappa = m$ . It is a consequence of the axioms of a  $\otimes$ -algebra that a *shuffle-idempotent* is an *idempotent*.

**Theorem 1.** *Let  $L \subseteq \Sigma^{\otimes}$  be a regular language. The following are equivalent.*

1.  $L$  is recognized by some finite *shuffle-trivial*  $\otimes$ -algebra.
2.  $L$  is a boolean combination of languages of the form  $B^{\otimes}$  where  $B \subseteq \Sigma$ .
3.  $L$  is definable in  $\text{FO}^1$ .
4.  $L$  is *recognized* by direct product of  $\mathbf{U}_1$ s.
5. The syntactic  $\otimes$ -algebra of  $L$  is *shuffle-trivial*.

*Proof.*

(1  $\Rightarrow$  2) Let  $L$  be *recognized* by  $h: \Sigma^{\otimes} \rightarrow \mathbf{M}$  where  $\mathbf{M} = (M, \text{id}, \cdot, \tau, \tau^*, \kappa)$  is a *shuffle-trivial*  $\otimes$ -algebra. Note that, by  $\otimes$ -algebra axioms, for any  $m \in M$ , we have  $m^\kappa \cdot m^\kappa = m^\kappa$ . If  $m$  is a *shuffle-idempotent*, that is if  $m^\kappa = m$ , we get that  $m = m \cdot m$ . Thus a *shuffle-idempotent* is necessarily an *idempotent*. So  $\mathbf{M}$  is a *commutative, shuffle-idempotent* (and hence *idempotent*)  $\otimes$ -algebra, meaning its every element is a *shuffle-idempotent* (and hence an *idempotent*).

Consider an arbitrary *word*  $u \in \Sigma^{\otimes}$  and let  $\text{alpha}(u) \subseteq \Sigma$  be the set of *letters* in the word  $u$ . We show that  $h(u) = h(a_1) \cdot \dots \cdot h(a_n)$  where  $\{a_1, \dots, a_n\} = \text{alpha}(u)$ . For the empty word, it is trivially true. For any non-empty word, the

proof uses [evaluation trees](#) introduced in Section 2. Let  $\mathcal{T} = (T, h)$  be an [evaluation tree](#) over  $u$ . We show by induction on the tree that  $h(v) = h(a_1) \cdot \dots \cdot h(a_n)$  where  $\{a_1, \dots, a_n\} = \text{alpha}(v)$  for all nodes  $v$  in the tree. Consider a node  $v$  of the tree.

1. Case  $v$  is a letter: The induction hypothesis clearly holds.
2. Case  $v$  is a concatenation of words  $v_1$  and  $v_2$ : This is same as in the classical finite word case. The induction hypothesis holds since it holds for both  $v_1$  and  $v_2$ , and since all elements of  $\mathbf{M}$  are [idempotents](#), and [commutative](#).
3. Case  $v$  is an omega sequence of words  $\langle v_1, v_2, \dots \rangle$  such that there exists an  $e \in M$  and  $h(v_i) = e$  for all  $i \geq 1$  and  $h(v) = e^\tau$ . Since from the axioms of  [\$\otimes\$ -algebra](#)  $e^\kappa = (e^\kappa)^\tau$  we have  $h(v) = e$ . Clearly there is a  $k \geq 1$  such that  $\text{alpha}(v_1 v_2 \dots v_k) = \text{alpha}(v)$  and therefore it suffices to show that  $h(v_1 v_2 \dots v_k) = h(v)$ . This is true, since  $e$  is an idempotent  $h(v_1 v_2 \dots v_k) = h(v_1) \cdot h(v_2) \cdot \dots \cdot h(v_k) = e$ .
4. Case  $v$  is an omega\* sequence of words: This is symmetric to the case above. The induction hypothesis follows from the following axiom of  [\$\otimes\$ -monoid](#):  $e^\kappa = (e^\kappa)^{\tau^*}$
5. Case  $v$  is a [perfect shuffle](#) such that  $h(v) = \{b_1, \dots, b_k\}^\kappa$ . By the shuffle-trivial property, we have  $h(v) = b_1 \cdot \dots \cdot b_k$ . Let  $v = \prod_{i \in \mathbb{Q}} v_i$  where  $h(v_i) \in \{b_1, \dots, b_k\}$ . By induction hypothesis  $h(v_i) = h(a_1^i) \cdot \dots \cdot h(a_n^i)$  where  $\text{alpha}(v_i) = \{a_1^i, \dots, a_n^i\}$ . Let  $l \geq k$  and  $j_1, j_2, \dots, j_l \in \mathbb{Q}$  be such that we get the following:  $\{h(v_{j_1}), h(v_{j_2}), \dots, h(v_{j_l})\} = \{b_1, \dots, b_k\}$  and  $\text{alpha}(v_{j_1} \dots v_{j_l}) = \text{alpha}(v)$ . Let  $w = v_{j_1} \dots v_{j_l}$ . Since elements of  $\mathbf{M}$  are [commutative](#) and [idempotents](#),  $h(v) = h(w) = h(v_{j_1}) \cdot \dots \cdot h(v_{j_l})$ . This shows that the induction hypothesis also holds in this case, as it reduces to the finite concatenation case.

The induction hypothesis, therefore, holds for any word  $u \in A^\otimes$ . So  $L$  is union of equivalence classes defined by the finite index relation  $\{(u, v) \mid \text{alpha}(u) = \text{alpha}(v)\}$ . All these classes are boolean combination of [languages](#) of the form  $B^\otimes$  for some  $B \subseteq \Sigma$ , as seen below.

$$\{u \mid \text{alpha}(u) = B\} = B^\otimes \setminus \left( \bigcup_{b \in B} (B \setminus \{b\})^\otimes \right)$$

(2  $\Rightarrow$  3) Note that  $B^\otimes$  is expressed by the  $\text{FO}^1$  formula  $\forall x \bigvee_{a \in B} a(x)$ . The claim follows from boolean closure of  $\text{FO}^1$ .

(3  $\Rightarrow$  4) Due to the restriction of a single variable, any formula  $\phi(x)$  is a boolean combination of atomic [letter](#) predicates. Since a position in a word can have exactly one [letter](#), any non-trivial formula  $\phi(x)$  is a disjunction of [letter](#) predicates, e.g.  $a(x) \vee b(x)$ . A [language](#) defined by the sentence  $\exists x (a(x) \vee b(x))$  is recognized by the  [\$\otimes\$ -monoid](#)  $\mathbf{U}_1$  via  $h: \Sigma \rightarrow \mathbf{U}_1$  that maps  $a, b$  to  $0 \in \mathbf{U}_1$  and every other letter to  $\text{id} \in \mathbf{U}_1$ . A [language](#) defined by boolean combination of such sentences can be [recognized](#) by direct products of  $\mathbf{U}_1$ .

(4  $\Rightarrow$  5) The syntactic  [\$\otimes\$ -monoid](#) of  $L$  [divides](#) any  [\$\otimes\$ -monoid](#) that [recognizes](#)  $L$ ; so it [divides](#) a direct product of finitely many  $\mathbf{U}_1$ . It is almost trivially verified



that  $\otimes$ -monoid  $U_1$  is commutative and shuffle-trivial. Since these properties are equivalent to a set of identities and identities are preserved under direct product and division, we get that the syntactic  $\otimes$ -monoid of  $L$  is commutative and shuffle-trivial.

(5  $\Rightarrow$  1) The syntactic  $\otimes$ -algebra of  $L$  is finite because  $L$  is a regular language. Also, it is commutative and shuffle-trivial by assumption, and a language is always recognized by its syntactic  $\otimes$ -algebra. So this direction trivially holds.  $\square$

We point out an interesting connection of the above result to the block product based characterizations from [1]. As shown there,  $FO^3$  (resp.  $FO^2$ ) are characterized by strongly (resp. weakly) iterated block products of copies of  $U_1$ . In the same spirit,  $FO^1$  is characterized by direct-products of copies of  $U_1$ .

### 3.2 Boolean closure of existential FO

Let us first recall the characterization of  $B(\exists^*)$  - the boolean closure of existential FO over finite words. This is precisely the content of the theorem due to Simon [14]. The usual presentation of Simon's theorem refers to *piecewise testable languages* which are easily seen to be equivalent to  $B(\exists^*)$ -definable languages. *Simon's theorem* states that a regular language of finite words is  $B(\exists^*)$ -definable iff its syntactic monoid is *J-trivial*. We recall that a monoid  $M$  is *J-trivial* if and only if for all  $m, n \in M$ ,  $MmM = MnM$  implies  $m = n$ . In short, the Green's equivalence relation  $J$  on  $M$  is the equality relation. We refer to [11] for a detailed study of Green's relations and its use in the proof of Simon's theorem.

The original proof of Simon's theorem uses the congruence  $\sim_n$ , parametrized by  $n \in \mathbb{N}$ , on finite words  $\Sigma^*$ : for  $u, v \in \Sigma^*$ ,  $u \sim_n v$  if  $u$  and  $v$  have the same set of subwords of length less than or equal to  $n$ . Note that  $\sim_n$  has finite index. It turns out that the finite quotient monoid  $\Sigma^* / \sim_n$  is *J-trivial*. Furthermore, every finite *J-trivial* monoid is a quotient of the  $\Sigma^* / \sim_m$  for an appropriate choice of  $\Sigma$  and  $m \in \mathbb{N}$ . It is known [17] that a finite monoid  $(M, \cdot)$  is *J-trivial* iff it satisfies the (profinite) identities<sup>3</sup>:  $x^! = x \cdot x^!$  and  $(x \cdot y)^! = (y \cdot x)^!$ . See the proof of the following theorem in [11, Theorem 3.13].

**Theorem 2 (Simon's theorem [11]).** *Let  $\mathbf{M} = (M, \text{id}, \cdot)$  be a *J-trivial* monoid. Consider the morphism  $h : \Sigma^* \mapsto \mathbf{M}$ . Then there exists an  $n$  such that for all  $u, v \in \Sigma^*$ ,  $u \sim_n v$  implies  $h(u) = h(v)$ .*

We fix  $n \in \mathbb{N}$  and work with  $\sim_n$  defined on countable words  $\Sigma^{\otimes}$ : for  $u, v \in \Sigma^{\otimes}$ ,  $u \sim_n v$  if  $u$  and  $v$  have the same set of subwords of length less than or equal to  $n$ . It is immediate that  $\sim_n$  is an equivalence relation on  $\Sigma^{\otimes}$  of finite index. We let  $S_n = \Sigma^{\otimes} / \sim_n$  denote the finite set of  $\sim_n$ -equivalence classes. For a word  $w$ ,  $[w]_n$  denotes the  $\sim_n$ -equivalence class which contains  $w$ .

**Lemma 1.** *There is a natural well-defined product operation  $\pi : S_n^{\otimes} \rightarrow S_n$  as follows:  $\pi\left(\prod_{i \in \alpha} [w_i]_n\right) = \left[\prod_{i \in \alpha} w_i\right]_n$ . This operation  $\pi$  satisfies the generalized associativity property. As a result,  $\mathbf{S}_n = (S_n, \text{id} = [\varepsilon]_n, \pi)$  is a  $\otimes$ -monoid.*

<sup>3</sup> We denote the unique idempotent power of  $m$  by  $m^!$

*Proof.* Let  $w = \prod_{i \in \alpha} w_i$  and  $w' = \prod_{i \in \alpha} w'_i$  where  $w_i \sim_n w'_i$  for all  $i \in \alpha$ . To prove  $\pi$  is well-defined, we show that  $\pi(w) = \pi(w')$ . It suffices to show that all subwords of  $w$  of length less than or equal to  $n \in \mathbb{N}$  are also subwords of  $w'$ .

Consider an arbitrary subword  $u$  of  $w$  that is of length less than or equal to  $n$ . If  $u$  is the empty subword, it also is a subword of  $w'$ . Otherwise, we can break  $u$  into non-empty factors  $u = u_1 u_2 \dots u_k$  where  $k \leq n$  and  $u_j$  (for  $1 \leq j \leq k$ ) is a subword of  $w_{i_j}$  ( $i_j < i_{j'}$  whenever  $j < j'$ ). Since  $w_{i_j} \sim_n w'_{i_j}$  and  $|u_j| \leq |u| \leq n$ , we have  $u$  is a subword of  $w'$  as well. Therefore,  $\pi$  is well-defined.

Now we turn to the issue of proving [generalized associativity](#) property of  $\pi$ . Let  $u = \prod_{i \in \alpha} u_i$  where  $u_i = \prod_{j \in \alpha_i} [v_j]_n$  and  $\alpha$  is any countable [linear ordering](#). We have  $\pi(u_i) = [\prod_{j \in \alpha_i} v_j]_n$  and hence

$$\pi\left(\prod_{i \in \alpha} \pi(u_i)\right) = \left[\prod_{i \in \alpha} \left(\prod_{j \in \alpha_i} v_j\right)\right]_n = \pi(u)$$

This completes the proof.  $\square$

Note that the lemma implies that  $h_n : \Sigma^{\otimes} \rightarrow \mathbf{S}_n$  mapping  $w$  to  $[w]_n$  is a [morphism](#) of  $\otimes$ -monoids.

**Lemma 2.** *Every countable word  $u$  has a finite subword  $\hat{u}$  such that  $u \sim_n \hat{u}$ .*

*Proof.* Let  $u$  be an arbitrary [countable word](#) and let  $W$  be the set of [subwords](#) of length  $n$  or less in  $u$ . For a  $v \in W$  identify one set  $X_v \subseteq \text{dom}(u)$  such that  $v = u|_{X_v}$ . Let  $X = \cup_{v \in W} X_v$  and  $\hat{u} = u|_X$  the [subword](#) corresponding to  $u$  restricted to the finite set of positions in  $X$ . Then,  $u \sim_n \hat{u}$ .  $\square$

We say that a  $\otimes$ -algebra is [shuffle-power-trivial](#) if it satisfies the (profinite) [identity](#):

$$\{x_1, \dots, x_p\}^{\kappa} = (x_1 \cdot x_2 \cdot \dots \cdot x_p)^{\dagger}$$

Note that, every [idempotent](#) of such a  $\otimes$ -algebra is a [shuffle-idempotent](#):  $x^{\dagger} = x$  implies  $x^{\kappa} = x$ . Moreover it satisfies the [identities](#) for  $J$ -trivial:  $(x \cdot y)^{\dagger} = \{x, y\}^{\kappa} = \{y, x\}^{\kappa} = (y \cdot x)^{\dagger}$  and  $x^{\dagger} = x \cdot x^{\dagger}$  follows from the axioms of  $\otimes$ -algebra and  $x^{\dagger} = x^{\kappa} = (x^{\kappa})^{\kappa} = (x^{\kappa})^{\tau} = (x^{\dagger})^{\tau} = x^{\tau} = x \cdot x^{\tau}$ . It follows that [shuffle-power-trivial  \$\otimes\$ -algebras](#) are aperiodic.

**Theorem 3.** *Let  $L \subseteq \Sigma^{\otimes}$  be a regular language. The following are equivalent.*

1.  $L$  is [recognized](#) by a finite [shuffle-power-trivial  \$\otimes\$ -algebra](#).
2.  $L$  is [recognized](#) by the quotient [morphism](#)  $h_n : \Sigma^{\otimes} \rightarrow \mathbf{S}_n$  for some  $n$ .
3.  $L$  is [definable](#) in  $B(\exists^*)$ .
4. The syntactic  $\otimes$ -algebra of  $L$  is [shuffle-power-trivial](#).

*Proof.*

(1  $\Rightarrow$  2) Let  $L$  be recognized by  $h : \Sigma^{\otimes} \rightarrow \mathbf{M}$  where  $\mathbf{M} = (M, \text{id}, \cdot, \tau, \tau^*, \kappa)$  is a [shuffle-power-trivial  \$\otimes\$ -algebra](#). Since the [identities](#) are preserved in sub- $\otimes$ -algebra, we can assume  $h$  to be surjective. Consider the restriction of  $h$  to the

free monoid  $\Sigma^*$  resulting in the induced monoid morphism, also denoted  $h$  by slight abuse of notation,  $h: \Sigma^* \rightarrow (M, \text{id}, \cdot)$ . By the **identities** of the  $\otimes$ -*algebra*  $\mathbf{M}$ , this morphism is surjective.

By Simon's theorem, there exists  $n \in \mathbb{N}$ , such that  $(M, \text{id}, \cdot)$  is a quotient of  $\Sigma^*/\sim_n$  and  $u \sim_n v$  implies  $h(u) = h(v)$  thereby creating the quotienting morphism defined by mapping  $[u]_n \mapsto h(u)$ . This implies in our morphism  $h: \Sigma^\otimes \rightarrow \mathbf{M}$ , for all finite words  $u, v$ , we have  $u \sim_n v$  implies  $h(u) = h(v)$ . Note that by Lemma 2, for every countable word  $w$  there exists a finite subword of it,  $\hat{w}$  such that  $w \sim_n \hat{w}$ . We now show that  $h(w) = h(\hat{w})$ . In the remainder of this proof,  $\hat{u}$  for a countable word  $u$  denotes a subword of  $u$  such that  $u \sim_n \hat{u}$  where existence of  $\hat{u}$  is ensured by Lemma 2.

If  $w$  is the empty word, then  $\hat{w}$  is also the empty word, and the property holds. Otherwise, let  $\mathcal{T} = (T, h)$  be an **evaluation tree** over  $w$ . We prove by induction on the tree that for every node  $v$  of the tree,  $h(v) = h(\hat{v})$ .

1. Case  $v$  is a letter: The induction hypothesis clearly holds by taking  $\hat{v} = v$ .
2. Case  $v$  is a concatenation of words  $v_1$  and  $v_2$ : Note that  $\hat{v}_1 \sim_n v_1$  and  $\hat{v}_2 \sim_n v_2$  implies  $\hat{v}_1 \hat{v}_2 \sim_n v_1 v_2$  and also  $\hat{v}_1 \hat{v}_2$  is a finite subword of  $v_1 v_2$ . By induction hypothesis,  $h(v_1) = h(\hat{v}_1)$  and  $h(v_2) = h(\hat{v}_2)$ . Hence  $h(v_1 v_2) = h(\hat{v}_1 \hat{v}_2)$ . This proves the induction hypothesis holds in this case.
3. Case  $v$  is an omega sequence of words  $\langle v_1, v_2, \dots \rangle$  such that there exists an idempotent  $e \in M$  and  $h(v_i) = e$  for all  $i \geq 1$  and  $h(v) = e^\tau$ . Since from the axioms of  $\otimes$ -*algebra*  $e^\kappa = (e^\kappa)^\tau$  we have  $h(v) = e$ . Because there are only finitely many words of length less than or equal to  $n$ , clearly there is a  $k \geq 1$  such that  $v_1 v_2 \dots v_k \sim_n v$ . Let us denote  $v_1 v_2 \dots v_k$  by  $v'$ . Note that since  $e$  is an idempotent,  $h(v') = e = h(v)$ . Also by the induction hypothesis and the concatenation case already seen above,  $h(v') = h(\hat{v}')$ . By transitivity,  $\hat{v}'$  is a finite subword of  $v$  that is  $\sim_n$  equivalent to it, and  $h(v) = h(v') = h(\hat{v}')$ . This proves the induction hypothesis for this case.
4. Case  $v$  is an omega\* sequence of words: This is symmetric to the case above. The induction hypothesis follows from the following axiom of  $\otimes$ -*monoid*:  $e^\kappa = (e^\kappa)^\tau^*$
5. Case  $v$  is a **perfect shuffle** such that  $h(v) = \{b_1, \dots, b_k\}^\kappa$ . By the **shuffle-power-trivial** property, we have  $h(v) = (b_1 \dots b_k)^\dagger$ . Let  $v = \prod_{i \in \mathbb{Q}} v_i$  where  $h(v_i) \in \{b_1, \dots, b_k\}$ . By induction hypothesis  $h(v_i) = h(\hat{v}_i)$ . Since there are only finitely many words of length less than or equal to  $n$ , there exists  $l \geq k$  and  $j_1, j_2, \dots, j_l \in \mathbb{Q}$  such that we get the following:  $\{h(v_{j_1}), h(v_{j_2}), \dots, h(v_{j_l})\} = \{b_1, \dots, b_k\}$  and  $v_{j_1} \dots v_{j_l} \sim_n v$ . Let  $w = v_{j_1} \dots v_{j_l}$ . Let  $w^m$  be a finite power of  $w$  such that  $h(w^m) = h(w)^\dagger$ . Since the  $\otimes$ -*monoid* satisfies the **J-trivial** identity  $(x \cdot y)^\dagger = (y \cdot x)^\dagger$ , we have that  $h(w^m) = h(v)$  and  $w^m$  is a finite subword of  $v$  that is  $\sim_n$ -equivalent to  $v$ . This shows that the induction hypothesis also holds in this case.

Now for any two countable words  $u$  and  $v$ , if  $u \sim_n v$ , then  $h(u) = h(\hat{u}) = h(\hat{v}) = h(v)$  where the middle equality is from the classical result of Simon mentioned before. Having shown that  $u \sim_n v$  implies  $h(u) = h(v)$ , it now immediately

follows that  $L$  is a boolean combination of  $\sim_n$  equivalence classes and hence is recognized by a morphism to  $\Sigma^\otimes/\sim_n$ .

(2  $\Rightarrow$  1) Let  $[x]_n J [y]_n$  for some  $x, y \in \Sigma^\otimes$ . There exists  $[u]_n, [v]_n, [u']_n, [v']_n$  such that  $[u]_n \cdot [x]_n \cdot [v]_n = [y]_n$  and  $[u']_n \cdot [y]_n \cdot [v']_n = [x]_n$ . In other words,  $uxv \sim_n y$  and  $u'yv' \sim_n x$ . This implies  $x \sim_n y$ , that is,  $[x]_n = [y]_n$ . This proves  $\Sigma^\otimes/\sim_n$  is  $J$ -trivial.

It is not difficult to see that  $\{u_1, \dots, u_p\}^n \sim_n (u_1 u_2 \dots u_p)^n$ . This means  $\{[u_1]_n, \dots, [u_p]_n\}^\kappa = ([u_1]_n \dots [u_p]_n)^\dagger$ .

(2  $\Rightarrow$  3) Every equivalence class of  $\sim_n$  is clearly definable in  $B(\exists^*)$ .

(3  $\Rightarrow$  2) Let  $L$  be recognized by the formula  $\alpha ::= \exists x_1, \dots, x_n \varphi(x_1, \dots, x_n)$ . We show that for an  $u \sim_n v$ ,  $u \models \alpha$  if and only if  $v \models \alpha$ . Consider an assignment  $s$  which assigns the variables  $x_i$ s to a position in the domain of  $u$  such that  $u, s \models \varphi$ . Note that since  $\varphi$  is a quantifier free formula it is a boolean combination of formulas of the form  $x_i < x_j$ ,  $x_i = x_j$  and  $a(x_i)$ . Let  $X = \{s(x_i) \mid 1 \leq i \leq n\} \subseteq \text{dom}(u)$  be the set of  $n$  points which are assigned to the  $x_i$ s. Since  $u \sim_n v$ , there is a set  $Y \subseteq \text{dom}(v)$  of  $n$  points such that  $u|_X = v|_Y$ . Consider an assignment  $\hat{s}$  to variables  $x_i$  to positions in  $Y$  such that  $s(x_i) < s(x_j)$  iff  $\hat{s}(x_i) < \hat{s}(x_j)$ . Clearly such an assignment satisfies  $v, \hat{s} \models \varphi$  since the ordering between the variables and the letter positions are preserved. Therefore we get that  $u \models \alpha$  implies  $v \models \alpha$ . A symmetric argument shows that other direction too.

(4  $\Rightarrow$  1) This is a trivial observation.

(1  $\Rightarrow$  4) This follows from the fact that [identities](#) are preserved under [division](#).  $\square$

### 3.3 Summary of FO subclasses

We summarize the results of this section and known results from the literature.

Logic	Identity	Decomposition
FO <sup>1</sup>	shuffle-trivial	$\times(U_1)$
$B(\exists^*)$	shuffle-power-trivial	$\mathbf{S}_n$
FO <sup>2</sup>	$\otimes$ -DA [9]	$wbp(U_1)$ [1]
FO	$x^\tau \cdot x^{\tau^*} = x$ and shuffle simple [5]	$bp(U_1)$ [1]

## 4 First Order Logic with infinitary quantifiers

Our results in the previous section resemble very closely the corresponding results over finite words. This can be attributed to the limited capability of the operators  $\tau$ ,  $\tau^*$  and  $\kappa$  in the  $\otimes$ -monoids we witnessed. As mentioned in the Introduction, FO cannot define the language of infinite number of  $a$ 's. An existential quantifier is a threshold counting quantifier - it says there exists at least one position satisfying a property. Using multiple such first-order quantifiers, FO can count

up to any finite constant but not more. Over **countable words**, it is natural to ask for stronger threshold quantifiers. We introduce natural infinite extensions of the existential quantifier. These quantifiers can distinguish ordinals in the infinite.

We define  $\mathcal{I}_0$  to be the set of all non-empty finite orderings. For any number  $n \in \mathbb{N}$ , we define the set  $\mathcal{I}_n$  to be the set of all **orderings** of the form  $\sum_{i \in \mathbb{Z}} \alpha_i$  where  $\alpha_i \in \mathcal{I}_{n-1} \cup \{\varepsilon\}$  and is closed under finite sum. We define the **Infinitary rank** (or simply **rank**) of a **linear ordering**  $\alpha$  (denoted by  $\mathcal{I}\text{-rank}(\alpha)$ ) as the least  $n$  (if it exists) where  $\alpha \in \mathcal{I}_n$ . If there is no such  $n$  we say that the rank is infinite. For example,  $\mathcal{I}\text{-rank}(\omega) = \mathcal{I}\text{-rank}(\omega + \omega) = \mathcal{I}\text{-rank}(\omega^* + \omega) = 1$ ,  $\mathcal{I}\text{-rank}(\omega^2) = \mathcal{I}\text{-rank}(\omega^2 + \omega^*) = 2$ , and the **rank** of rational numbers is infinite.

We introduce the logic  $\text{FO}[\infty]$  extending  $\text{FO}$  with **infinitary quantifiers** :  $\exists^{\infty_0} x \varphi \mid \exists^{\infty_1} x \varphi \mid \dots \mid \exists^{\infty_n} x \varphi \mid \dots$  for all  $n \in \mathbb{N}$ .

Note that all the variables are first order. The semantics of the **infinitary quantifier**  $\exists^{\infty_n} x \varphi$  for an  $n \geq 0$  is: for a word  $w$  and an assignment  $s$ , we say  $w, s \models \exists^{\infty_n} x \varphi$  if there exists a **subordering**  $X \subseteq \text{dom}(w)$  such that  $\mathcal{I}\text{-rank}(X) = n$  and  $w, s[x = i] \models \varphi$  for all  $i \in X$ . For example,  $\exists^{\infty_0} x \varphi$  is equivalent to  $\exists x \varphi$  since both formulas are true if and only if there is at least one satisfying assignment  $x = s$ .

The logic  $\text{FO}[(\infty_j)_{j \leq n}]$  denote the fragment containing only the infinitary quantifiers  $\exists^{\infty_j} x$  for all  $j \leq n$ . Clearly the following relationship is maintained among the logics:

$$\text{FO} = \text{FO}[(\infty_j)_{j \leq 0}] \subseteq \text{FO}[(\infty_j)_{j \leq 1}] \subseteq \text{FO}[(\infty_j)_{j \leq 2}] \subseteq \dots$$

We also denote by  $\text{FO}^1[(\infty_j)_{j \leq n}]$  the corresponding one variable fragment of  $\text{FO}[(\infty_j)_{j \leq n}]$ .

*Example 3.* The formula  $\exists^{\infty_1} x a(x)$  denotes the set of all countable words with infinitely many positions labelled  $a$ . Since  $\text{FO}$  cannot express this, it shows  $\text{FO} \subsetneq \text{FO}[(\infty_j)_{j \leq 1}]$ .

For an  $n \geq 0$ , we define the  $\otimes$ -algebra  $\Delta_n$ -chains as:  $(\{\text{id}, 0, 1, \dots, n\}, \text{id}, \cdot, \tau, \tau^*, \kappa)$  where for all  $0 \leq i \leq j \leq n$ ,  $i \cdot j = j \cdot i = \max(i, j) = j$  and for all  $0 \leq k < n$ ,  $k^\tau = k^{\tau^*} = k + 1$  and  $n^\tau = n^{\tau^*} = n$ . That is,  $k^\tau = k^{\tau^*} = \min(k + 1, n)$ . Moreover,  $\text{id}^\kappa = \text{id}$  and  $S^\kappa = n$  for any  $S$  where  $S \setminus \{\text{id}\} \neq \emptyset$ .

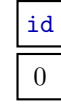
$$\Delta_n ::= (\{\text{id}, 0, 1, \dots, n\}, \{i, j\} \mapsto \max(i, j), i \mapsto \min(i+1, n), i \mapsto \min(i+1, n), S \mapsto n)^4$$

Note that the syntactic  $\otimes$ -algebra for the language defined by  $\exists^{\infty_n} x a(x)$  is  $\Delta_n$ .

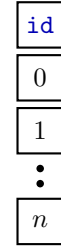
#### 4.1 $\text{FO}[\infty]$ with single variable

In this section we show that languages recognized by  $\Delta_n$  are definable in  $\text{FO}^1[(\infty_j)_{j \leq n}]$ . It is easy to observe that the **direct product** of  $\Delta_n$  recognize exactly those **languages** definable in the one variable fragment.

<sup>4</sup> As mentioned in the Preliminaries, we restrict the descriptions of **derived operators** to  $\Delta_n \setminus \{\text{id}\}$



$\Delta_0$ -chain



**Theorem 4.** *Languages recognized by direct product of  $\Delta_n$  are exactly those definable in  $\text{FO}^1[(\infty_j)_{j \leq n}]$ .*

*Proof.* We first show that languages recognized by  $\Delta_n$  are definable in  $\text{FO}^1[(\infty_j)_{j \leq n}]$ .

Let  $h: \Sigma^{\otimes} \rightarrow \Delta_n$  be a morphism. It suffices to show that for any element  $m \in \Delta_n$ ,  $h^{-1}(m)$  is definable in  $\text{FO}^1[(\infty_j)_{j \leq n}]$ . In the rest of the discussion we adopt the convention that  $\text{id} < 0$ . Let  $\uparrow m$  denote the set  $\{m' \mid m' \geq m\}$ . Note that for an  $m < n$ ,  $h^{-1}(m) = h^{-1}(\uparrow m) \setminus h^{-1}(\uparrow(m+1))$  and  $h^{-1}(n) = h^{-1}(\uparrow n)$ . Therefore, it is sufficient to show that  $h^{-1}(\uparrow m)$  is definable in  $\text{FO}^1[(\infty_j)_{j \leq n}]$ . For each  $m \in \Delta_n$ , we define the language  $L(m)$  as

$\{w \mid \text{there exists a letter } a \text{ in } w \text{ such that } h(a) = j \neq \text{id} \text{ and either } j \geq m \text{ or there is a set of positions } \alpha \text{ labelled } a \text{ such that } \mathcal{I}\text{-rank}(\alpha) = j' \text{ and } j + j' \geq m\}$   
The following  $\text{FO}^1[(\infty_j)_{j \leq n}]$  sentence defines the language  $L(m)$ .

$$\bigvee_{a \in \Sigma, h(a) \geq m} \exists x a(x) \quad \vee \quad \bigvee_{a \in \Sigma, 0 \leq h(a) < m} \exists^{\infty_{m-h(a)}} x a(x)$$

We show that  $L(m) = h^{-1}(\uparrow m)$  by induction on  $m$ . The base case holds since  $\uparrow \text{id} = \Delta_n$ ,  $h^{-1}(\uparrow \text{id}) = \Sigma^{\otimes}$  and  $L(\text{id}) = \Sigma^{\otimes}$ . To prove the induction hypothesis assume the claim holds for all  $j < m$ . Consider a word  $w$ . By a second induction on the height of an evaluation tree  $(T, h)$  for  $w$  we show for all words  $v \in T$ ,  $v \in h^{-1}(\uparrow m)$  if and only if  $v \in L(m)$ . In each of the following cases we assume that the children of the node (if they exist) satisfy the second induction hypothesis.

1. Case  $v$  is a letter: The hypothesis clearly holds
2. Case  $v$  is a concatenation of two words  $v_1$  and  $v_2$ : There are two cases to consider -  $\{v_1, v_2\} \cap h^{-1}(\uparrow m) \neq \emptyset$  or not. In the first case, let for an  $i \in \{1, 2\}$  we have  $h(v_i) \geq m$  and  $v_i \in L(m)$ . Clearly  $h(v) = h(v_1 v_2) \geq m$  and  $v \in L(m)$ . For the second case, let us assume  $h(v_1) = i$  and  $h(v_2) = j$  such that  $i \leq j < m$  and both  $v_1, v_2 \notin L(m)$ . From the definition of  $\Delta_n$ , it follows that  $h(v) = h(v_1 v_2) = j$ . Let the  $a$ -labelled suborderings in  $v_1$  and  $v_2$  be  $\alpha_1$  and  $\alpha_2$  respectively where  $\mathcal{I}\text{-rank}(\alpha_1) \leq \mathcal{I}\text{-rank}(\alpha_2) = j'$ . It follows from the definition that  $\mathcal{I}\text{-rank}(\alpha_1 + \alpha_2) = j'$  and therefore  $v \notin L(m)$ .
3. Case  $v$  is an omega sequence of words  $\langle v_1, v_2, \dots \rangle$  such that  $h(v_i) = k$ , for all  $i$ , and  $k$  is an idempotent (in  $\Delta_n$  all elements are idempotents): Firstly, if  $k \geq m$  and  $v_i \in L(m)$  then clearly  $h(v) \geq m$  and  $v \in L(m)$ . The non-trivial case is  $k = m - 1$ . From the second induction hypothesis  $v_i \notin L(m)$  for all  $i$ . From the definition of  $\Delta_n$ ,  $h(v) = k^\tau = m$ . We need to show that  $v \in L(m)$ . By first induction hypothesis, each  $v_i$  has a letter  $a_i$  and an  $a_i$ -labelled set of positions  $\alpha_i$  such that  $h(a_i) + \mathcal{I}\text{-rank}(\alpha_i) = k$ . Since  $|\Sigma|$  is finite, there is a letter  $a$  occurring in omega many factors. Hence the  $a$ -labelled set of positions  $\alpha$  in  $v$  satisfies  $h(a) + \mathcal{I}\text{-rank}(\alpha) = k + 1$  or in other words  $v \in L(m)$ .
4. Case  $v$  is an omega\* sequence: This case is symmetric to the above case.
5. Case  $v$  is a perfect shuffle,  $h(v) = S^\kappa$ : It is easy to see that the induction hypothesis holds if  $S = \{\text{id}\}$ . So, assume  $S \cap \{\text{id}\} \neq \emptyset$ . Hence  $h(v) = n$ . Since, there are rational number of children  $u$  where  $h(u) \neq \text{id}$ , there is a letter  $a$  such that  $a$ -labelled set of positions in  $v$  has infinite rank or  $v \in L(n)$ .

Now we give the proof of the other direction. Due to the restriction of a single variable, any quantifier free formula  $\varphi(x)$  is a boolean combination of atomic letter predicates. Since any position has exactly one **letter**, we can consider  $\varphi(x)$  to be a disjunction of the **letter** predicates. Consider the formula  $\alpha ::= \exists^{\infty k} x \bigvee_{a \in A} a(x)$ . The language defined by  $\alpha$  is **recognized** by the morphism  $h : \Sigma^{\otimes} \mapsto \Delta_n$  where  $h(a) = 0$  for all  $a \in A$  and  $h(a) = \mathbf{id}$  for all  $a \notin A$  since  $h^{-1}(\{k, k+1, \dots, n\})$  **recognizes** all words where the  $A$ -labelled positions have **rank** at least  $k$ . We conclude by observing that **languages** defined by boolean combinations of such sentences can be **recognized** by **direct products** of  $\Delta_n$ .  $\square$

## 4.2 The general $\mathbf{FO}[\infty]$ logic

In this section, we consider the full logic  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  and observe that they define exactly those **languages recognized** by block products of  $\Delta_n$ .

**Theorem 5.** *The languages defined by  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  are exactly those **recognized** by finite **block products** of  $\Delta_n$ . Moreover, the **languages** defined by  $\mathbf{FO}[\infty]$  are exactly those **recognized** by finite **block products** of  $\{\Delta_n \mid n \in \mathbb{N}\}$ .*

*Proof.* We first show that **languages recognizable** by finite **block products** of  $\Delta_n$  are definable in  $\mathbf{FO}[(\infty_j)_{j \leq n}]$ . The proof is via induction on the number of  $\Delta_n$  in an iterated **block product**. The base case follows from Theorem 4.

For the inductive step, consider a **morphism**  $h : \Sigma^{\otimes} \rightarrow M \square \Delta_n$ . Let  $h_1 : \Sigma^{\otimes} \rightarrow M$  be the induced morphism to  $M$ , and let  $\sigma$  be the associated transducer. By the **block product principle** (see Proposition 1), any language **recognized** by  $h$  is a boolean combination of **languages**  $L_1 \subseteq \Sigma^{\otimes}$  **recognized** by  $M$  and  $\sigma^{-1}(L_2)$  where  $L_2 \subseteq (M \times \Sigma \times M)^{\otimes}$  is **recognized** by  $\Delta_n$ . By induction hypothesis,  $L_1$  is  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  definable. By the base case  $L_2$  is  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  definable but over the alphabet  $M \times \Sigma \times M$ . To complete the proof, one needs to show for any word  $w \in \Sigma^{\otimes}$  and assignment  $s$ , and for any  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  formula  $\varphi$  over the alphabet  $M \times \Sigma \times M$ , there exists a  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  formula  $\hat{\varphi}$  over the alphabet  $\Sigma$  such that  $w, s \models \hat{\varphi}$  if and only if  $\sigma(w), s \models \varphi$ . For instance, suppose  $\varphi = \exists^{\infty i} x (m_1, c, m_2)(x)$ , and inductively  $\phi_{m_1}$  (resp.  $\phi_{m_2}$ ) are  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  formula characterizing **words** over  $\Sigma^{\otimes}$  that are mapped by  $h_1$  to  $m_1$  (resp.  $m_2$ ). Then  $\hat{\varphi}$  is  $\exists^{\infty i} x (\phi_{m_1}|_{<x} \wedge c(x) \wedge \phi_{m_2}|_{>x})$ , where  $\phi_{m_1}|_{<x}$  is the formula  $\phi_{m_1}$  with all its variables relativised to less than the variable  $x$ . This way, one proves that  $\sigma^{-1}(L_2)$  is  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  definable. This completes the proof of this direction.

We now show the other direction: a **language** defined by an arbitrary formula  $\varphi \in \mathbf{FO}[(\infty_j)_{j \leq n}]$  is **recognized** by finite **block products** of  $\Delta_n$ . For any  $\mathbf{FO}[(\infty_j)_{j \leq n}]$  formula, we can naturally consider its models  $w, s$  as words over extended alphabets; for each free variable an 1 or 0 is added to the label of a position in  $w$  depending on whether  $s$  assigns the corresponding free variable to that position or not. The proof now goes via structure induction on the subformulas  $\phi$  of  $\varphi$ .

Case  $\phi = a(x)$ : Then  $L(\phi) \subseteq (A \times \{0, 1\})^{\otimes}$  is the set of **words** with a unique position labeled  $(a, 1)$ . This can be **recognized** by  $U_1 \square U_1 [1]$ . Since  $U_1$  divides  $\Delta_n$ , the hypothesis holds.

Case  $\phi = x < y$ : This can also be **recognized** by **block products** of  $U_1 [1]$ .

Boolean combination of formula can be handled by **direct product** of inductively defined  $\otimes$ -algebras.

Case  $\phi = \exists^{\infty i} x \psi$  (for  $i \leq n$ ): This is the non-trivial case. Let  $L(\psi) \subseteq (\Sigma \times \{0, 1\})^{\otimes}$  is inductively **recognized** by  $h: (\Sigma \times \{0, 1\})^{\otimes} \rightarrow M \in \square \Delta_n$ , that is, there is a set  $F \subseteq M$  such that  $h^{-1}(F) = L(\psi)$ . We prove that  $M \square \Delta_n$  recognizes  $L(\phi)$ . Once again we use the **block product principle**. Consider two **morphisms**  $g_1: \Sigma^{\otimes} \rightarrow M$  and  $g_2: (M \times \Sigma \times M)^{\otimes} \rightarrow \Delta_n$ . Let  $g_1(a) = h((a, 0))$  and suppose  $g_2((m_1, a, m_2))$  equals 0 if  $m_1 \cdot h((a, 1)) \cdot m_2 \in F$ , and it equals **id** otherwise. Let  $\sigma$  be the transducer corresponding to  $g_1$ . We show that  $w \models \phi$  if and only if  $g_2(\sigma(w)) \geq i$ . This would imply  $L(\phi) = \sigma^{-1}(g_2^{-1}(\{i, i+1, \dots, n\}))$  and by the **block product principle**, this is recognized by  $M \square \Delta_n$ .

Let  $w \models \phi$ . If  $\alpha_\psi$  is the set of all positions of  $w$  where  $\psi$  is true, then  $\mathcal{I}\text{-rank}(w_\psi) \geq i$ . Let  $l \in \alpha_\psi$  and  $w(l) = a$ . We can split  $w$  at the position  $l$  as  $w_1 a w_2$  and by logic semantics  $w_1^0(a, 1) w_2^0 \models \psi$  (for any  $u \in \Sigma^{\otimes}$ , we denote by  $u^0$  the word over the same domain with  $u^0[i] = (u[i], 0)$ ). If  $h(w_1^0) = m_1$  and  $h(w_2^0) = m_2$ , then  $m_1 \cdot h((a, 1)) \cdot m_2 \in F$ . Also,  $\sigma(w)[l] = (m_1, a, m_2)$ . So,  $g_2$  maps every position  $l \in \alpha_\psi$  to 0, and hence  $g_2(\sigma(w)) \geq i$ . Conversely, suppose  $g_2(\sigma(w)) \geq i$ . Let  $\alpha_0$  denote the positions of  $\sigma(w)$  for which  $g_2$  maps to 0. Since  $g_2$  maps each letter to 0 or **id**, we get  $\mathcal{I}\text{-rank}(\alpha_0) \geq i$ . Let  $l \in \alpha_0$ . If  $\sigma(w)(l) = (m_1, a, m_2)$ , then  $m_1 \cdot h((a, 1)) \cdot m_2 \in F$ . This means  $\psi$  is true at position  $l$  for  $w$ . Since  $l$  is any position in  $\alpha_0$ , we have that  $w \models \phi$ .  $\square$

We claim that both first order logic with **cuts** (**FO[cut]**) and weak monadic second order logic (**WMSO**) can define the languages definable in **FO[ $\infty$ ]**.

**Theorem 6.** **FO[ $\infty$ ]  $\subseteq$  FO[cut]  $\cap$  WMSO** <sup>5</sup>

*Proof.* We first show by structural induction that there is an equivalent **WMSO** formula for any **FO[ $\infty$ ]** formula. It is easy to observe that the hypothesis holds for the atomic case, first order quantification and boolean combinations. Let us consider the formula  $\phi = \exists^{\infty k} x \psi(x)$ . By our inductive hypothesis there is a **WMSO** formula  $\hat{\psi}(x)$  equivalent to  $\psi(x)$ . We show that the **WMSO** formula  $\Psi_k$  inductively defined is equivalent to  $\phi$ : Let  $\Psi_0 ::= \exists x \hat{\psi}(x)$  and

$\Psi_n ::=$  “For any finite set  $X = \{x_1, \dots, x_k\}$ , one of the **intervals**  $[-, x_1], \dots, [x_i, x_{i+1}], \dots, [x_k, -]$  can be split into at least two parts each satisfying  $\Psi_{n-1}$ ”

This can be expressed in **WMSO** as follows (**consec**( $X, x, y$ ) says that  $x, y \in X$  and  $x < y$  and there is no  $z \in X$  such that  $x < z < y$ ). That is  $x$  and  $y$  are

<sup>5</sup> Here, **FO[ $\infty$ ]**, **FO[cut]**, **WMSO** denote the languages defined by the respective logic.



consecutive in set  $X$ ):

$$\Psi_n ::= \forall_{fin} X \left( \exists x, y \in X \exists z (\text{consec}(X, x, y) \rightarrow \Psi_{n-1}[> x, < z] \wedge \Psi_{n-1}[> z, < y]) \vee \right. \\ \left. \exists z (\Psi_{n-1}[> z, < \min(X)] \wedge \Psi_{n-1}[< z]) \vee \exists z (\Psi_{n-1}[> \max(X), < z] \wedge \Psi_{n-1}[> z]) \right)$$

We claim that  $\Psi_n$  satisfies all words where the  $\psi$ -labelled set of positions  $\alpha$  has  $\mathcal{I}\text{-rank}(\alpha) \geq n$ . It is clearly true for the base case  $\Psi_0$ . Assume the hypothesis is true for all  $j < n$ . The formula  $\Psi_n$  says that for any finite number of partitions  $\alpha_1, \alpha_2, \dots, \alpha_k$  of the  $\psi$ -labelled set of positions  $\alpha$ , there is at least one  $\alpha_i$  that can be split into two parts containing  $\psi$ -labelled set of positions  $\alpha_i^1$  and  $\alpha_i^2$  such that  $\mathcal{I}\text{-rank}(\alpha_i^1) \geq n - 1$  and  $\mathcal{I}\text{-rank}(\alpha_i^2) \geq n - 1$ . In short, finite partitioning of  $\psi$ -labelled set of positions with rank  $n - 1$  is not possible or  $\mathcal{I}\text{-rank}(\alpha) \geq n$ . Therefore the formula  $\Psi_k$  is equivalent to the formula  $\phi$ .

Next we give an **FO[cut]** formula equivalent to an **FO[ $\infty$ ]** formula. Like in the previous proof, let us look at the case  $\phi = \exists^{\infty} x \psi(x)$  where  $\psi(x)$  is equivalent to an **FO[cut]** formula  $\hat{\psi}(x)$ . We show  $\phi$  is equivalent to  $\Phi_k$  where  $\Phi_n$  is inductively defined as:  $\Phi_0 ::= \exists x \hat{\psi}(x)$  and  $\Phi_n$  is

“There is a **cut** towards which there is an **omega** (or **omega\***) sequence of **factors** satisfying  $\Phi_{n-1}$ ”

This can be written in **FO[cut]** as follows:

$$\exists_{cut} \hat{x} \forall y < \hat{x} \exists z \in (y, \hat{x}) \Phi_{n-1}[> y, < z] \vee \exists_{cut} \hat{x} \forall y > \hat{x} \exists z \in (\hat{x}, y) \Phi_{n-1}[> z, < y]$$

The formula says there is an **omega** or **omega\*** sequence of  $\psi$ -labelled positions which are of **rank**  $n - 1$ . We now argue that the inductively defined formulas  $\Phi_n$  says there is a  $\psi$ -labelled set of positions  $\alpha$  where  $\mathcal{I}\text{-rank}(\alpha) \geq n$ . The base is same as the previous case. Let us consider the formula  $\Phi_n$ . We argue that the first (resp. second) disjunct says there is an **omega** (resp. **omega\***) sequence of **domains**  $\alpha_1, \alpha_2, \dots$ , such that each of the  $\alpha_i$ 's are  $\psi$ -labelled positions and  $\mathcal{I}\text{-rank}(\alpha_i) \geq n - 1$ . The first (resp. second) disjunct guesses a **cut** towards which this **omega** (resp. **omega\***) sequence approaches. It then says that for any point strictly before (resp. after) this **cut**, there is a point strictly after (resp. before) between which there is a  $\psi$ -labelled set of positions  $\alpha$  where  $\mathcal{I}\text{-rank}(\alpha) \geq n - 1$ . Clearly, this implies the  $\psi$ -labelled set of positions have **rank**  $\geq n$ . We conclude by observing that the **FO[cut]** formula  $\Phi_k$  is equivalent to the formula  $\phi$ .  $\square$

## 5 No Finite Basis Theorems

The main goal of this section is to prove that **FO[ $\infty$ ]**, **FO[cut]** and **FO[cut]  $\cap$  WMSO** over **countable words** do not admit a **block product** based characterization which uses only a **finite** set of  **$\oplus$ -monoids**. This is in stark contrast with the result in [1] which shows that a language of **countable words** is **FO**-definable iff it is recognized by a strong iteration of **block product** of copies of  $\Delta_0$  (alternately called  $U_1$ ). This is abbreviated by saying that **FO** has a block-product based

characterization using a basis which contains the single  $\otimes$ -monoid  $\Delta_0$ . Notice that, it follows from the results in the previous section that  $\text{FO}[\infty]$  admits a **block product** based characterization using the *natural infinite basis*  $\{\Delta_n\}_{n \in \mathbb{N}}$ .

Fix a finite  $\otimes$ -algebra  $\mathbf{M} = (M, \text{id}, \cdot, \tau, \tau^*, \kappa)$ . For every  $n \in \mathbb{N}$ , we define the operation  $\gamma_n : M \rightarrow M$  which maps  $x$  to  $x^{\gamma_n}$ . The inductive definition of  $\gamma_n$  is as follows:  $x^{\gamma_0} = x^!$  and  $x^{\gamma_n} = ((x^{\gamma_{n-1}})^{\tau} (x^{\gamma_{n-1}})^{\tau^*})^!$ .

**Lemma 3.** *For each  $m \in M$ , there exists  $n$  such that  $\forall n' \geq n, m^{\gamma_n} = m^{\gamma_{n'}}$ .*

*Proof.* Let  $m$  be an arbitrary element in a  $\otimes$ -monoid  $\mathbf{M}$ . We show that there exists an  $n$  such that for all  $n' > n$  we have  $m^{\gamma_n} = m^{\gamma_{n'}}$ . Consider the following sequence:  $a_0 = m$  and  $a_{j+1} = ((a_j)^{\tau} \cdot (a_j)^{\tau^*})^!$ . Note that  $m^{\gamma_j} = a_j$  for all  $j$ . We argue that for consecutive  $a_j, a_{j+1}$  either  $a_j >_{\mathcal{J}} a_{j+1}$  or  $a_{j+1} = a_j$ . Let us assume  $a_j \mathcal{J} a_{j+1}$ . Clearly  $a_{j+1}$  is  $\mathcal{R}$  and  $\mathcal{L}$  equivalent to  $a_j$ . Therefore  $a_j \mathcal{H} a_{j+1}$ . Since in a  $\mathcal{J}$  class containing a group  $a_j \mathcal{J} a_j^{\tau}$  we have that the  $\mathcal{H}$  class of  $a_j$  has cardinality one and therefore  $a_j = a_{j+1}$ .  $\square$

We now define the *gap-nesting-length* of  $\mathbf{M}$  (in notation,  $\text{gnlen}(\mathbf{M})$ ) to be the smallest  $n$  such that for all  $m \in M$ ,  $m^{\gamma_n} = m^{\gamma_{n+1}}$ . It follows from the previous lemma that a finite  $\otimes$ -algebra has a finite *gap-nesting-length*. It is a simple computation that, for each  $k$ ,  $\text{gnlen}(\Delta_k) = k$ . The following main technical lemma is the key to our no-finite-basis theorems.

**Lemma 4.** *For finite aperiodic<sup>6</sup>  $\otimes$ -algebras  $\mathbf{M}$  and  $\mathbf{N}$ ,*

1. *We have,  $\text{gnlen}(\mathbf{M} \square \mathbf{N}) \leq \max(\text{gnlen}(\mathbf{M}), \text{gnlen}(\mathbf{N}))$ .*
2. *If  $\mathbf{M}$  divides  $\mathbf{N}$  then  $\text{gnlen}(\mathbf{M}) \leq \text{gnlen}(\mathbf{N})$ .*

Before proving Lemma 4, we first recall the product and omega operations of semidirect products from [1]. Consider two  $\otimes$ -algebras  $\mathbf{M} = (M, \text{id}, \cdot, \tau, \tau^*, \kappa)$  and  $\mathbf{N} = (N, \widehat{\text{id}}, +, \widehat{\tau}, \widehat{\tau}^*, \widehat{\kappa})$ . We denote the left and right action of  $M$  on  $N$  by  $*$ . The semidirect product is  $\mathbf{M} \times \mathbf{N} = (M \times N, \tilde{\cdot}, \tilde{\tau}, \tilde{\tau}^*, \tilde{\kappa})$  where the operations are suitably defined. We recall the product operation and the omega and omega\* operation for an element  $(e, n)$  where  $e$  is an idempotent.

1.  $(m_1, n_1) \tilde{\cdot} (m_2, n_2) = (m_1 \cdot m_2, n_1 * m_2 + m_1 * n_2)$
2.  $(e, n)^{\tilde{\tau}} = (e^{\tau}, n * e^{\tau} + (e * n * e^{\tau})^{\widehat{\tau}})$
3.  $(e, n)^{\tilde{\tau}^*} = (e^{\tau^*}, (e^{\tau^*} * n * e)^{\widehat{\tau}^*} + e^{\tau^*} * n)$

The block product of  $\mathbf{M}$  and  $\mathbf{N}$ , denoted  $\mathbf{M} \square \mathbf{N}$ , is the semidirect product of  $\mathbf{M}$  and  $\mathbf{K}$  where  $\mathbf{K}$  is the direct product of  $|M| \times |M|$  copies of  $\mathbf{N}$ . Note that the underlying set of  $\mathbf{K}$  can also be considered as  $N^{M \times M}$ : the set of all functions from  $M \times M$  to  $N$ . We refer to [1] for action axioms over  $\otimes$ -algebra.

**Lemma 5.** *Let  $(m, f) \in \mathbf{M} \square \mathbf{N}$  be an idempotent. Then  $m \in M$  is an idempotent and  $m * f * m$  is an idempotent in  $N^{M \times M}$ .*

<sup>6</sup> This means the underlying monoid of a  $\otimes$ -algebra is aperiodic

*Proof.* Since  $(m, f)$  is an idempotent, by concatenation rule of block product algebra, we have  $(m, f) = (m^2, f * m + m * f)$ . Hence  $m = m^2$ , that is,  $m \in M$  is an idempotent. Also,  $f = f * m + m * f$  implies  $m * f * m = m * f * m + m * f * m$ . Hence,  $m * f * m$  is an idempotent in  $N^{M \times M}$ .  $\square$

**Lemma 6.** *Let  $(m, f), (m', f') \in \mathbf{M} \square \mathbf{N}$  such that  $(m, f) = (m', f')^!$ . Then  $m = (m')^!$ . Further, if  $M$  is aperiodic, then  $m * f * m = (m * f' * m)^!$ .*

*Proof.* Suppose  $k$  is an idempotent power of  $(m', f')$ . So  $m = (m')^k$  and  $f = \sum_{i=0}^{k-1} (m')^i * f' * (m')^{k-i-1}$ . By Lemma 5,  $m$  is an idempotent, so  $m = (m')^!$ . If  $M$  is aperiodic, then  $(m')^j = m$  for  $j \geq k$ . Hence  $m * f * m = (m * f' * m)^k$ . By Lemma 5,  $m * f * m$  is an idempotent. So  $m * f * m = (m * f' * m)^!$ .  $\square$

**Lemma 7.** *Consider  $\otimes$ -algebra  $\mathbf{M}$  has compatible left and right actions on  $\otimes$ -algebra  $\mathbf{P}$ . Let  $m, m' \in M$  and  $p \in P$ . Then  $m * p^{\gamma_n} * m' = (m * p * m')^{\gamma_n}$*

*Proof.* We first prove that  $m * p^! * m' = (m * p * m')^!$ . By action axioms for concatenation, it is easy to see that  $m * p^k * m' = (m * p * m')^k$  for any natural number  $k \geq 1$ . Note that there exists  $k \in \mathbb{N}$  such that  $p^k = p^!$  and  $(m * p * m')^k = (m * p * m')^!$ . Then  $m * p^! * m' = m * p^k * m' = (m * p * m')^k = (m * p * m')^!$ .

The proof is now by induction on  $n$ . For  $n = 0$ , we have  $m * p^{\gamma_0} * m = m * p^! * m = (m * p * m)^! = (m * p * m)^{\gamma_0}$ .

For the inductive step, note that

$$\begin{aligned}
 m * p^{\gamma_n} * m' &= m * ((p^{\gamma_{n-1}})^{\tau} \cdot (p^{\gamma_{n-1}})^{\tau^*})^! * m' && \text{defn. of } \gamma_n \\
 &= (m * ((p^{\gamma_{n-1}})^{\tau} \cdot (p^{\gamma_{n-1}})^{\tau^*}) * m')^! \\
 &= ((m * (p^{\gamma_{n-1}})^{\tau} * m') \cdot (m * (p^{\gamma_{n-1}})^{\tau^*} * m'))^! && \text{action axiom for } \cdot \\
 &= ((m * (p^{\gamma_{n-1}}) * m')^{\tau} \cdot (m * (p^{\gamma_{n-1}}) * m')^{\tau^*})^! && \text{action axiom for } \tau, \tau^* \\
 &= (((m * p * m')^{\gamma_{n-1}})^{\tau} \cdot ((m * p * m')^{\gamma_{n-1}})^{\tau^*})^! && \text{induction hypothesis} \\
 &= (m * p * m')^{\gamma_n} && \text{defn. of } \gamma_n
 \end{aligned}$$

This completes the proof.  $\square$

**Lemma 8.** *Consider  $(m, f), (m', f') \in \mathbf{M} \square \mathbf{N}$  such that  $(m, f) = (m', f')^{\gamma_n}$ . Then  $m = (m')^{\gamma_n}$ . If  $\mathbf{M}$  is aperiodic, then  $m * f * m = (m * f' * m)^{\gamma_n}$ .*

*Proof.* The proof is by induction on  $n$ . For the base case of  $n = 0$ , we have  $(m, f) = (m', f')^{\gamma_0} = (m', f')^!$ . By Lemma 6,  $m = (m')^! = m^{\gamma_0}$  and if  $\mathbf{M}$  is aperiodic,  $m * f * m = (m * f' * m)^! = (m * f' * m)^{\gamma_0}$ . This proves the base case.

For the inductive step, let  $(m, f) = (m', f')^{\gamma_n} = ((m', f')^{\gamma_{n-1}})^{\gamma_1}$ . Also let  $(e, g) = (m', f')^{\gamma_{n-1}}$ . So  $(m, f) = (e, g)^{\gamma_1}$ . By induction hypothesis,  $e = (m')^{\gamma_{n-1}}$  and  $m = e^{\gamma_1}$  implying  $m = ((m')^{\gamma_{n-1}})^{\gamma_1} = (m')^{\gamma_n}$ . If  $\mathbf{M}$  is aperiodic, then by induction hypothesis,  $e * g * e = (e * f' * e)^{\gamma_{n-1}}$  and  $m * f * m = (m * g * m)^{\gamma_1}$ .

Note that since  $m = e^{\gamma_1} = (e^{\tau} \cdot e^{\tau^*})^!$ , we have  $m \cdot e = e \cdot m = m$ . Therefore

$$\begin{aligned}
m * f * m &= (m * g * m)^{\gamma_1} \\
&= (m * (e * g * e) * m)^{\gamma_1} \\
&= (m * (e * f' * e)^{\gamma_{n-1}} * m)^{\gamma_1} \\
&= ((m * f' * m)^{\gamma_{n-1}})^{\gamma_1} \\
&= (m * f' * m)^{\gamma_n}
\end{aligned}$$

This completes the proof.  $\square$

*Proof (of Lemma 4).*

We first prove the first case. Consider two aperiodic  $\otimes$ -monoids  $\mathbf{M}$  and  $\mathbf{N}$  such that  $\max(\text{gnlen}(\mathbf{M}), \text{gnlen}(\mathbf{N})) = k \in \mathbb{N}$ . We show that  $\text{gnlen}(\mathbf{M} \square \mathbf{N}) \leq k$ . Note that, for any  $m \in \mathbf{M}$  and any  $n \in \mathbf{N}$ ,  $m^{\gamma_k} = m^{\gamma_{k+1}}$  and  $n^{\gamma_k} = n^{\gamma_{k+1}}$ .

Let  $(m, f) \in M \square N$  be an arbitrary element. We show that  $(m, f)^{\gamma_k} = (m, f)^{\gamma_{k+1}}$ . Let  $(e, g) = (m, f)^{\gamma_k}$ . Then  $(e, g)^{\gamma_1} = (m, f)^{\gamma_{k+1}}$ . Also by Lemma 8,  $e = m^{\gamma_k}$  and  $e * g * e = (e * f * e)^{\gamma_k}$ . Since  $\mathbf{M}$  and  $\mathbf{N}$  have gap-nesting-length less than or equal to  $k$ , we get  $e = m^{\gamma_k} = m^{\gamma_{k+1}} = e^{\gamma_1}$  and  $e * g * e = (e * f * e)^{\gamma_k} = (e * f * e)^{\gamma_{k+1}} = (e * g * e)^{\gamma_1}$ . Now we use the fact that in any  $\otimes$ -algebra  $x = x^{\gamma_1}$  implies  $xJx^{\tau}$  and  $xJx^{\tau^*}$  and that further implies  $x = x^{\tau} \cdot x^{\tau^*}$ . See [5] for the details regarding the simple proof of this property based on Green's relations.

Therefore we have  $e = e^{\tau} \cdot e^{\tau^*}$  and  $e * g * e = (e * g * e)^{\tau} + (e * g * e)^{\tau^*}$ . Since  $(e, g)$  is an idempotent by definition of the  $\gamma_i$  operation, we get that  $e$  is an idempotent in  $\mathbf{M}$  by Lemma 5. Therefore

$$\begin{aligned}
&(e, g)^{\tau} \cdot (e, g)^{\tau^*} \\
&= (e^{\tau} e^{\tau^*}, g * e^{\tau} e^{\tau^*} + (e * g * e^{\tau} e^{\tau^*})^{\tau} + (e^{\tau} e^{\tau^*} * g * e)^{\tau^*} + e^{\tau} e^{\tau^*} * g) \\
&= (e, g * e + (e * g * e)^{\tau} + (e * g * e)^{\tau^*} + e * g) \\
&= (e, g * e + e * g * e + e * g) \\
&= (e, g)^3 \\
&= (e, g)
\end{aligned}$$

Hence  $(m, f)^{\gamma_{k+1}} = (e, g)^{\gamma_1} = (e, g) = (m, f)^{\gamma_k}$ . This completes the proof for the block product operation.

Now we prove the second case. If  $\mathbf{M}$  is a subalgebra of  $\mathbf{N}$ , then the property is easily verified. Let's suppose  $h: \mathbf{N} \rightarrow \mathbf{M}$  is a surjective morphism, and  $\text{gnlen}(\mathbf{N}) = k$ . For any  $m \in \mathbf{M}$ , there exists  $n \in \mathbf{N}$  such that  $h(n) = m$ . It is straightforward to check that  $m^{\gamma_k} = h(n^{\gamma_k}) = h(n^{\gamma_{k+1}}) = m^{\gamma_{k+1}}$ . This completes the proof.  $\square$

**Corollary 1.**  $\text{FO}[(\infty_j)_{j \leq n}] \subsetneq \text{FO}[(\infty_j)_{j \leq n+1}]$ .

*Proof.* By Theorem 5, the syntactic  $\otimes$ -algebra  $\mathbf{M}$  of any  $\text{FO}[(\infty_j)_{j \leq n}]$ -definable language divides a block product of copies of  $\Delta_n$ . By Lemma 4 and the fact that  $\text{gnlen}(\Delta_k) = n$ ,  $\text{gnlen}(\mathbf{M}) \leq n$ . Note that,  $\Delta_{n+1}$  is the syntactic  $\otimes$ -algebra

for the language  $L$  defined by the  $\text{FO}[(\infty_j)_{j \leq n+1}]$  formula  $\exists^{\infty_{n+1}} x a(x)$ . As  $\text{gnlen}(\Delta_{n+1}) = n + 1$ , it follows that  $L$  cannot be defined in  $\text{FO}[(\infty_j)_{j \leq n}]$ .  $\square$

**Theorem 7.** *There is no finite basis for a block product based characterization for any of these logical systems  $\text{FO}[\infty]$ ,  $\text{FO}[\text{cut}]$ ,  $\text{FO}[\text{cut}] \cap \text{WMSO}$ .*

*Proof.* Fix one of the logics  $\mathcal{L}$  mentioned in the statement of the theorem. It follows from Theorem 6 and the algebraic characterization [5] of  $\text{FO}[\text{cut}]$  that the syntactic  $\otimes$ -algebras of  $\mathcal{L}$ -definable languages are aperiodic. Now suppose, for contradiction, let  $\mathcal{L}$  admit a finite basis  $B$  of aperiodic  $\otimes$ -algebras for its block product based characterization. Since  $B$  is finite, there exists  $n \in \mathbb{N}$  such that for all  $\otimes$ -algebras  $\mathbf{M}$  in  $B$ ,  $\text{gnlen}(\mathbf{M}) \leq n$ . It follows by Lemma 4 that the syntactic  $\otimes$ -algebra  $\mathbf{N}$  of every  $\mathcal{L}$ -definable language has the property  $\text{gnlen}(\mathbf{N}) \leq n$ .

Now consider the language  $L$  defined by the  $\text{FO}[\infty]$  sentence  $\phi = \exists^{\infty_{n+1}} x a(x)$ . By Theorem 6,  $L$  is  $\mathcal{L}$ -definable. Hence, the gap-nesting-length of the syntactic  $\otimes$ -algebra  $K$  of  $L$  is less than or equal to  $n$ . However,  $K$  is simply  $\Delta_{n+1}$  and  $\text{gnlen}(\Delta_{n+1}) = n + 1$ . This leads to a contradiction.  $\square$

## References

1. Adsul, B., Sarkar, S., Sreejith, A.V.: Block products for algebras over countable words and applications to logic. In: 34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019, Vancouver, BC, Canada, June 24-27, 2019. pp. 1–13. IEEE (2019). <https://doi.org/10.1109/LICS.2019.8785669>, <https://doi.org/10.1109/LICS.2019.8785669>
2. Baudisch, A., Seese, D., Tuschik, H.P., Weese, M.: Decidability and Generalized Quantifiers. Akademie Verlag, Berlin (1980)
3. Bès, A., Carton, O.: Algebraic characterization of FO for scattered linear orderings. In: Bezem, M. (ed.) Computer Science Logic, 25th International Workshop / 20th Annual Conference of the EACSL, CSL 2011, September 12-15, 2011, Bergen, Norway, Proceedings. LIPIcs, vol. 12, pp. 67–81. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2011)
4. Carton, O., Colcombet, T., Puppis, G.: An algebraic approach to MSO-definability on countable linear orderings. J. Symb. Log. **83**(3), 1147–1189 (2018). <https://doi.org/10.1017/jsl.2018.7>, <https://doi.org/10.1017/jsl.2018.7>
5. Colcombet, T., Sreejith, A.V.: Limited set quantifiers over countable linear orderings. In: Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Proceedings, Part II. pp. 146–158 (2015)
6. Diekert, V., Gastin, P., Kufleitner, M.: A survey on small fragments of first-order logic over finite words. Int. J. Found. Comput. Sci. **19**(3), 513–548 (2008). <https://doi.org/10.1142/S0129054108005802>, <https://doi.org/10.1142/S0129054108005802>
7. Gabbay, D.M., Hodkinson, I., Reynolds, M.: Temporal Logic: Mathematical Foundations and Computational Aspects, Volume 1. Oxford University Press, Oxford (1994)
8. Gradel, E., Otto, M., Rosen, E.: Two-variable logic with counting is decidable. In: Proceedings of Twelfth Annual IEEE Symposium on Logic in Computer Science. pp. 306–317 (1997). <https://doi.org/10.1109/LICS.1997.614957>

9. Manuel, A., Sreejith, A.V.: Two-variable logic over countable linear orderings. In: 41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016. pp. 66:1–66:13 (2016). <https://doi.org/10.4230/LIPIcs.MFCS.2016.66>, <https://doi.org/10.4230/LIPIcs.MFCS.2016.66>
10. Pin, J.É.: Handbook of formal languages, vol. 1. chap. Syntactic Semigroups, pp. 679–746. Springer-Verlag (1997), <http://dl.acm.org/citation.cfm?id=267846.267856>
11. Pin, J.É.: Mathematical foundations of automata theory (2020)
12. Rosenstein, J.G.: Linear orderings. Academic Press New York (1981)
13. Shelah, S.: The monadic theory of order. vol. 102, pp. 379–419 (1975)
14. Simon, I.: Piecewise testable events. In: Brakhage, H. (ed.) Automata Theory and Formal Languages. pp. 214–222. Springer Berlin Heidelberg (1975)
15. Straubing, H.: Finite automata, formal logic, and circuit complexity. Birkhauser Verlag, Basel, Switzerland (1994)
16. Straubing, H., Thérien, D., Thomas, W.: Regular languages defined with generalized quantifiers. In: Lepistö, T., Salomaa, A. (eds.) Automata, Languages and Programming. pp. 561–575. Springer Berlin Heidelberg, Berlin, Heidelberg (1988)
17. Straubing, H., Weil, P.: Varieties. CoRR [abs/1502.03951](https://arxiv.org/abs/1502.03951) (2015), <http://arxiv.org/abs/1502.03951>
18. Thomas, W.: Handbook of formal languages, vol. 3. chap. Languages, Automata, and Logic, pp. 389–455. Springer-Verlag New York, Inc. (1997), <http://dl.acm.org/citation.cfm?id=267871.267878>