Limited Set quantifiers over Countable Linear Orderings ⋆

Thomas Colcombet and A V Sreejith

LIAFA, Université Paris-Diderot
{thomas.colcombet, sreejith}@l iafa.univ-paris-diderot.fr

Abstract. In this paper, we study several sublogics of monadic second-order logic over countable linear orderings, such as first-order logic, first-order logic on cuts, weak monadic second-order logic, weak monadic second-order logic with cuts, as well as fragments of monadic second-order logic in which sets have to be well ordered or scattered. We give decidable algebraic characterizations of all these logics and compare their respective expressive power.

Key words: Linear orderings, Algebraic characterization, Monadic second order logic

1 Introduction

Monadic second-order logic (*i.e.*, first-order logic extended with set quantifiers) is a concise and expressive logic that retains good decidability properties (though with a bad complexity). In particular, since the seminal works of Büchi [3], Rabin [11] and Shelah [13], it is known to be decidable over infinite linear orderings with countably many elements, such as $(\mathbb{Q}, <)$ [5,7]. A breakthrough result of Shelah (also in [13]) states that over general linear orderings (*i.e.*, not necessarily countable), or simply over $(\mathbb{R}, <)$, this logic is not decidable anymore. There is also a long line of research focusing on the analysis of the expressive power and decidability status of temporal logics, which, for most of them are equivalent in expressiveness to first-order logic (but much more tractable), and can be decided on some non-countable linear orderings.

Such studies are interesting for themselves, *i.e.*, for the techniques involved in their resolution and the understanding of the logics it requires for doing so. Such studies are also interesting since infinite linear orderings offer a natural model of continuous linear time.

Recently, another step in our understanding of monadic second-order logic over countable linear orderings has been made. An algebraic model, $\circ$-monoids, was proposed [4], yielding among other results the first known quantifier collapse of monadic second-order logic (to the one alternation fragment over set quantifiers), the resolution of a conjecture of Gurevich and Rabinovich [8] concerning

⋆ The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n^o 259454.

the use of cuts "in the background" [6]. Algebraic recognizers give us a much deeper understanding of the expressive power of monadic second-order logic.

The next natural step is to follow the footprints of Schützenberger, who characterized algebraically first-order logic over finite words as languages that are recognized by aperiodic monoids [12] (in fact, the first-order logic terminology is in combination with McNaughton and Papert [10]) as these languages that are recognized by aperiodic monoids. Now that a suitable algebraic model is known for understanding monadic second-order logic, a similar study can be performed in this more general context. There exist already results of this kind, but these are so far restricted to the case of scattered linear orderings (*i.e.*, without any dense sub-ordering). In this context, first-order logic and first-order logic on cuts have been algebraically characterized [1], as well as weak monadic second-order logic [2]. Simple decision procedures are derived in all these situations.

In this paper, we perform a systematic analysis of sublogics of monadic second-order logic on countable linear orderings depending on the kind of sets over which set quantifiers range. If such sets are just singletons, we have exactly first-order logic (FO). If such sets are Dedekind cuts, we obtain first-order logic on cuts (FO[cut]). If finite sets only are allowed, this is weak monadic second-order logic (WMSO). If it is possible to quantify over both finite sets and cuts, we obtain weak monadic second-order on cuts (MSO[finite,cut]). We consider also MSO[ordinal] in which quantified sets need to be well-ordered. Finally MSO[scattered] corresponds to the case where quantified sets are required to be scattered. Our contribution is to compare the expressive power of all these logics (all are distinct but for MSO[finite,cut] which coincide with MSO[ordinal]), and characterize each of them by decidable algebraic means.

Structure of the paper In Section 2, we introduce linear orderings, words, and the logics we are interested in. In Section 3 we provide sufficient material concerning the algebraic framework of $\circ$-monoids, state the main characterization theorem, Theorem 2, and show the separation result, Theorem 3. Section 4 is devoted to the description of some ideas concerning one direction of the proof of Theorem 2. Section 5 concludes the paper.

2 Preliminaries

In this preliminary section, we introduce the notion of linear orderings (Section 2.1), (countable) words (Section 2.2) and the studied logics (Section 2.3).

2.1 Linear orderings

A *linear ordering* $\alpha = (X, <)$ is a non-empty set X equipped with a total order $<$. A linear ordering α is *dense* if it contains at least two elements and for all $x < y \in \alpha$, there exists a z such that $x < z < y$. It is *scattered* if no subset of X induces a dense ordering. A *well ordering* is a linear ordering such that every non-empty subset has a minimal element. A subset of a linear ordering is well ordered (*resp.* scattered) if the linear ordering restricted to it is a well ordering (*resp.* scattered).

Given an element x, its *successor* (resp. *predecessor*) (if it exists) is the only $y > x$ (*resp.* $y < x$) such that there is no z such that $x < z < y$ (*resp.* $y < z < x$). A subset $I \subseteq \alpha$ of a linear ordering is *convex* if whenever $x, y \in I$ and $x < z < y$, $z \in I$. A *condensation* of a linear ordering is an equivalence relation $\sim$ such that all equivalence classes are convex. For a linear ordering α and a condensation $\sim$, we denote by $\alpha/\sim$, the *condensed linear ordering*: its elements are the equivalence classes for $\sim$, and the ordering is obtained by projection of the original ordering. Two convex subsets I, J of a linear ordering are *consecutive* if I and J are disjoint and their union is convex. Using the notations for elements, if $I < J$, then I is the predecessor of J, while J is the successor of I.

Given linear orderings $(\beta_i)_{i\in\alpha}$ (assumed disjoint up to isomorphism) indexed with a linear ordering α, their generalized sum $\sum_{i\in\alpha} \beta_i$ is the linear ordering over the (disjoint) union of the sets of the β_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 β_i.

Given elements x, y, we denote by $[x, y)$ the set $\{z \mid x \leq z < y\}$, and similarly $[x, y]$, $(x, y]$ and (x, y). We also denote as $(-\infty, x)$, $(-\infty, x]$, $(x, +\infty)$ and $[x, +\infty)$ the intervals that are unlimited to the left or to the right. Usually Dedekind cuts are defined as ordered pairs of sets (L, R) such that $L < R$. Here, we define a *Dedekind cut* (or simply a cut) as a left-closed subset X of a linear ordering, *i.e.*, for all $x < y$ with $y \in X$, then $x \in X$.

2.2 Infinite words

Given a linear ordering α and a finite *alphabet* A, a *word* over A of *domain* α is a mapping $w : \alpha \to A$. The domain of a word is denoted $dom(w)$. In this work, all words are assumed of countable domain. The set of all *words of countable domain* is denoted by A°. A *language* is a subset of A°.

Given a convex set $X \subseteq dom(w)$ of word w, w_X denotes the word w *restricted* to X, *i.e.*, the word of domain X that coincides with w over X. A *factor* of a word w is any restriction of w to one of the convex subsets of its domain.

Given two words $u : \alpha \to A$ and $v : \beta \to A$ (where α and β are disjoint), we denote by uv the word over domain $\alpha + \beta$ such that each position $x \in \alpha$ (similarly $x \in \beta$) is labelled by $u(x)$ (by $v(x)$). The *generalized concatenation* of the words w_i (supposed of disjoint domain) indexed by a linear ordering α is

$$\prod_{i\in\alpha} w_i \ ,$$

and denotes the word of domain $\sum_i dom(w_i)$ which coincides with each w_i over $dom(w_i)$ for all $i \in \alpha$.

Some words will play an important role in the paper. The *empty word* ε, which is the only word of empty domain. The words denoted "$aaa\ldots$" and "$\ldots aaa$" are the words over the single letter a, and of respective domain $\omega = (\mathbb{N}, <)$ and $\omega^* = (\mathbb{N}, >)$. Finally, *perfectshuffle*(A) for A, a non-empty finite set of letters, is a word of domain $(\mathbb{Q}, <)$ in which all non-empty intervals (x, y) contain at least once each letter of A. This word is unique up to isomorphism.

2.3 First-order logic, monadic second-order logic, and between

We use logics for expressing properties of linear orderings or words. All of the several logics we study are all restrictions of monadic second-order logic (MSO). We very succinctly recall the basics of this logic here. The reader can refer to many other works on the subject, *e.g.*, [14]. We only consider word models.

Monadic second-order logic (MSO for short) is a logic with the following characteristics. It is possible to use *first-order variables* $x, y, z, \ldots$, ranging over positions of the word, and quantify over them thanks to $\exists x$ or $\forall y$. It is possible to use *monadic variables* $X, Y, \ldots$ (traditionally typeset in capital letters), that range over sets of positions of the word, and quantify over them using $\exists X$, $\forall Y$. Three atomic predicates can be used. The predicate $a(x)$, for a a letter, and x a position, holds if the letter carried at position x in the word is an a. The predicate $x < y$ for x, y, first-order variables denotes the order of the domain of the word. The membership predicate $x \in Y$ tests the membership of (the valuation of) a first-order variable x in (the valuation of) a monadic variable Y. All the Boolean connectives are also allowed. *First-order logic* (FO of short) is the fragment of this logic in which monadic variables, as well as quantifiers over them, are not allowed.

In this study, we are interested in the expressive power of logics weaker than MSO. There is a long tradition of such researches, initiated by the seminal work of Schützenberger. For instance, it is classical to study first-order logic and its fragments when the quantifier alternation or the number of variables are restricted. In our case, our goal is to investigate the intricate relationship between the expressive power of the logic, and the infinite/dense nature of the linear orderings/words under study. The only parameter that we use for modifying the power of the logic is to change the range of monadic variables. By default, such variables range over any set of positions. We introduce now several *restricted set quantifiers* and the corresponding logics. Our simplest logic is first-order logic. The logic obtained by allowing monadic quantifiers restricted to Dedekind cuts is denoted *FO[cut]*. Another situation is when monadic second-order variables range over finite set, yielding *weak monadic second-order logic* (WMSO for short). We are also interested in the fragment in which it is possible to quantify both over finite sets and Dedekind cuts. We denote this logic *MSO[finite,cut]*. Then come logics in which monadic variables range over "infinite but small", sets of positions. We consider the case in which it is possible to quantify over well ordered sets, or scattered sets. We denote these logics *MSO[ordinal]* and *MSO[scattered]*.

We formally denote these restricted quantifiers as $\exists^V$ and $\forall^V$, where $V \subseteq \{\in, \notin\}^\circ$. A set belongs to the range of the quantifier $\exists^V$ or $\forall^V$ if its characteristic map (as a labelling of the domain by $\in, \notin$) is in V.

Given one of the above logics $\mathcal{L}$, a formula $\varphi \in \mathcal{L}$ and a countable word w we denote by $w \models \varphi$, the fact that the formula is true over w. We say that w is a *model* of φ. A language $L \subseteq A^\circ$ is *definable* in $\mathcal{L}$ if there exists a formula φ in $\mathcal{L}$ such that for all words $w \in A^\circ$, $w \in L$ if and only if $w \models \varphi$.

Remark 1. Some dependencies between these logics are simple to establish:

FO[cut]

FO — MSO[finite,cut] — MSO[ordinal] – MSO[scattered] – MSO

WMSO

Indeed, FO[cut] is an extension of FO. Also WMSO extend FO since "being a singleton" is definable in WMSO. Similarly, MSO[finite,cut] is clearly an extension of both WMSO and FO[cut]. MSO[ordinal] can express finiteness, and represent cuts (as the left closure of a well ordered subset), and hence contains MSO[finite,cut]. In the same way, since being well ordered is expressible in MSO[scattered], MSO[scattered] contains MSO[ordinal]. Similarly, scatteredness being expressible in MSO, MSO[scattered] is a sublogic of MSO. In fact, all these logics are separated (Theorem 3), but for MSO[finite,cut] and MSO[ordinal] which happen to coincide (see Theorem 2).

The goal of this paper is to compare the expressive power of all these logics and be able to characterize them effectively.

3 The algebraic presentation: ∘-monoids

We now introduce the equivalent algebraic presentation of definable languages. We first describe the ∘-monoids in Section 3.1, and then the derived operations in Section 3.2, before presenting the theorems of characterization and separation in Section 3.3.

3.1 ∘-monoids, syntactic ∘-monoids and recognizability

As in the seminal work of Schützenberger, we use algebraic acceptors for describing regular languages of countable words: ∘-monoids. A ∘*-monoid* is a set M equipped with an operation π, called the *product*, from M° 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 α,

$$\pi\left(\prod_{i\in\alpha} u_i\right) = \pi\left(\prod_{i\in\alpha} \pi(u_i)\right) .$$

Of course, an instance of ∘-monoids is the set of words over some alphabet A equipped with the generalized concatenation $\prod$, *i.e.*, $(A^\circ, \prod)$. It is even the *free* ∘*-monoid* generated by A. A ∘*-monoid morphism* from $\mathbf{M}$ to $\mathbf{N}$ (∘-monoids) is a map γ from M to N such that $\gamma(\prod_{i\in\alpha} a_i) = \pi(\prod_{i\in\alpha} \gamma(a_i))$.

Example 1. **Sing**$= (\{1, s, 0\}, \pi)$ where π is defined for all $u \in \{1, s, 0\}^\circ$ as:

$$\pi(u) = \begin{cases} 1 & \text{if } u \in \{1\}^\circ, \\ s & \text{otherwise if } u \text{ contains no } 0\text{, and exactly one } s, \\ 0 & \text{otherwise,} \end{cases}$$

is a $\circ$-monoid (checking generalized associativity requires a case by case study).

By slightly modifying the example, we obtain the $\circ$-monoid **Fin** in which the second line in the definition of π is changed into "s if u contains no 0, and finitely many s's". The $\circ$-monoid **Ord** is when $\pi(u)$ evaluates to "s if u contains no 0, and a well ordered set of s's". Finally, the $\circ$-monoid **Scat** is when $\pi(u)$ evaluates to "s if u contains no 0, and a scattered set of s's". Once more, checking generalized associativity is by case analysis.

The element $\pi(\varepsilon)$ is called the *unit*, and it is customary to denote it 1 as done above. A *zero* (that does not necessarily exist) is an absorbing element, *i.e.*, an element such that $\pi(u0v) = 0$ whatever are u and v. It is denoted by convention 0 as in the above examples. An *idempotent* is an element e such that $\pi(ee) = e$.

A $\circ$-monoid can be used to recognize languages as follows. Consider a $\circ$-monoid $\mathbf{M} = (M, \pi)$, a map h from an alphabet A to M and a set $F \subseteq M$, then $(\mathbf{M}, h, F)$ *recognizes* the language $L = \{u \in A^\circ \mid \pi(h(u)) \in F\}$ (where h has been extended implicitly into a map from A° to M°). Said differently, L is the inverse image of F under the $\circ$-monoid morphism $\pi \circ h$. From [4], being recognizable by a $\circ$-monoid is equivalent to be definable in MSO.

Furthermore, when a language is recognizable by a finite $\circ$-monoid, then there is a minimal one called the *syntactic $\circ$-monoid*. It is minimal in the algebraic sense: all $\circ$-monoids that would recognize this language can be trimmed and quotiented yielding the syntactic one. We do not develop this aspect more in this short abstract.

Example 2. Coming back to the above examples, with $h(\in) = s$ and $h(\not\in) = 1$, then $(\mathbf{Sing}, h, \{s\})$ recognizes the language L_{Sing} over the alphabet $\{\in, \not\in\}$ of words that contain exactly one occurrence of $\in$. Similarly, $(\mathbf{Fin}, h, \{1, s\})$, $(\mathbf{Ord}, h, \{1, s\})$, and $(\mathbf{Scat}, h, \{1, s\})$ recognize the languages L_{Finite}, L_{Ord} and L_{Scat} respectively, of words that contain "finitely many $\in$'s", "a well ordered set of $\in$'s", and "a scattered set of $\in$'s" respectively.

Let us note that these languages are the one used in the restricted quantifiers $\exists^V$ and $\forall^V$ for defining the logics (cuts are omitted for space considerations).

3.2 The derived operations

The product operation π is infinite, even in a finite $\circ$-monoid $\mathbf{M} = (M, \pi)$. Hence, π is *a priori* not representable in finite space (it has uncountably many possible inputs). This problems is resolved using derived operations.

The *operations derived* from π are the following:

- 1 is the unit constant $\pi(\varepsilon)$,

– $\cdot$: $M \times M \to M$ is defined for $a, b \in M$ as $a \cdot b = \pi(ab)$,
– ω: $M \to M$ is defined for all $a \in M$ as $a^\omega = \pi(aaa\dots)$,
– $\omega*$: $M \to M$ is defined for all $a \in M$ as $a^{\omega*} = \pi(\dots aaa)$,
– η: $\mathcal{P}(M) \setminus \{\emptyset\} \to M$ is defined as $E^\eta = \pi(\mathit{perfectshuffle}(E))$ for $E \subseteq M$ non-empty.

Note that from the definitions, using generalized associativity, the unit element satisfies $1 \cdot 1 = 1^\omega = 1^{\omega*} = \{1\}^\eta = 1$, $a \cdot 1 = 1 \cdot a = a$, and $(E \cup \{1\})^\eta = E^\eta$ for all $a \in M$ and all non-empty $E \subseteq M$. Similarly, if there is a zero 0 then it satisfies $0 \cdot a = a \cdot 0 = 0^\omega = 0^{\omega*} = (E \cup \{0\})^\eta = 0$ for all $a \in M$ and $E \subseteq M$. This is why we usually do not mention these elements when describing derived operations.

Example 3. The derived operation of the above examples are entirely determined by the following table:

	$s \cdot s$	s^ω	$s^{\omega*}$	$\{s\}^\eta$
Sing	0	0	0	0
Fin	s	0	0	0
Ord	s	s	0	0
Scat	s	s	s	0

Though not essential in this short abstract, let us emphasize that the derived operations determine entirely the product π, as shown now.

Theorem 1. *There exists a set of equalities (A) involving the derived operations*[1]*, such that:*

– *The operations derived from a $\circ$-monoid satisfy all the equations from (A).*
– *If $1, \cdot, \omega, \omega^*, \eta$ are maps of correct type over a finite set M that satisfy the equalities of (A), then there exists one and only one product over M from which $1, \cdot, \omega, \omega^*, \eta$ are derived.*

3.3 The core theorem

We state in this section our main results, Theorem 2 and 3. All $\circ$-monoids are assumed finite from now. We first refine our understanding of idempotents:

– A *gap insensitive* idempotent e is an idempotent such that $e^\omega \cdot e^{\omega*} = e$.
– An *ordinal idempotent* e is an idempotent such that $e^\omega = e$. The name comes from the fact that in such a case, all words $u \in \{e\}^\circ$ that have a well ordered (*i.e.*, isomorphic to an ordinal) non-empty domain satisfy $\pi(u) = e$.
– Symmetrically, an *ordinal* idempotent* e is an idempotent such that $e^{\omega*} = e$.
– A *scattered idempotent* e is an idempotent which is at the same time an ordinal and an ordinal* idempotent. For such idempotents, all words $u \in \{e\}^\circ$ that have a scattered non-empty domain satisfy $\pi(u) = e$.

[1] These are variants of associativity, such as $x \cdot (y \cdot z) = (x \cdot y) \cdot z$, $1 \cdot x = x \cdot 1 = x$, $(a^n)^\omega = a^\omega$, and so on. A complete list is known [4], but of no use here.

– A *shuffle idempotent* e is an idempotent such that $\{e\}^\eta = e$.
– A shuffle idempotent e is *shuffle simple* if for all $K \subseteq M$ such that $e \cdot a \cdot e = e$ for all $a \in K$, $(\{e\} \cup K)^\eta = e$.

Note that since in every $\circ$-monoid $(\{e\}^\eta)^\omega = (\{e\}^\eta)^{\omega *} = \{e\}^\eta$, every shuffle idempotent is a scattered idempotent. Note also that every scattered idempotent is by definition an ordinal idempotent and an ordinal* idempotent. Also, every scattered idempotent is obviously gap insensitive.

We define now the following properties of a $\circ$-monoid $\mathbf{M} = (M, \pi)$:

– *aperiodic* if for all $a \in \mathbf{M}$ there exists n such that $a^n = a^{n+1}$,
– `i`$\to$`gi` if all idempotents are gap insensitive,
– `oi`$\to$`gi` if all ordinal idempotents are gap insensitive,
– `o`*`i`$\to$`gi` if all ordinal* idempotents are gap insensitive,
– `sc`$\to$`sh` if all scattered idempotents are shuffle idempotent,
– `sh`$\to$`ss` if all shuffle idempotents are shuffle simple.

It is clear by definition that `oi`$\to$`gi` (as well as `o`*`i`$\to$`gi`) imply `i`$\to$`gi`. There is in fact another, slightly less direct, implication to mention:

Lemma 1. `i`$\to$`gi` *implies aperiodic.*

Proof. Let a be an element of a finite $\circ$-monoid M. There exists n such that a^n is idempotent. We compute $a^n = (a^n)^\omega \cdot (a^n)^{\omega *} = a \cdot (a^n)^\omega \cdot (a^n)^{\omega *} = a^{n+1}$. □

We are now ready to state our core theorem.

Theorem 2. *Let* $\mathbf{M}$ *be the syntactic* $\circ$*-monoid of a language* $L \subseteq A^\circ$*, then:*

– *L is definable in FO iff* $\mathbf{M}$ *satisfies* `i`$\to$`gi`, `sc`$\to$`sh` *and* `sh`$\to$`ss`.
– *L is definable in FO[cut] iff* $\mathbf{M}$ *satisfies aperiodic,* `sc`$\to$`sh` *and* `sh`$\to$`ss`.
– *L is definable in WMSO iff* $\mathbf{M}$ *satisfies* `oi`$\to$`gi`, `o`*`i`$\to$`gi`, `sc`$\to$`sh` *and* `sh`$\to$`ss`.
– *L is definable in MSO[finite,cut] iff it is definable in MSO[ordinal] iff* $\mathbf{M}$ *satisfies* `sc`$\to$`sh` *and* `sh`$\to$`ss`.
– *L is definable in MSO[scattered] iff* $\mathbf{M}$ *satisfies* `sh`$\to$`ss`.

And as a consequence, these classes are decidable.

Example 4. Let us apply these characterizations to the $\circ$-monoids of Example 3:

	aper.	`i`$\to$`gi`	`oi`$\to$`gi`	`o`*`i`$\to$`gi`	`sc`$\to$`sh`	`sh`$\to$`ss`	definable in
Sing	yes	yes	yes	yes	yes	yes	FO
Fin	yes	no	yes	yes	yes	yes	WMSO, FO[cut], not FO
Ord	yes	no	no	yes	yes	yes	FO[cut], not WMSO
Scat	yes	yes	yes	yes	no	yes	MSO[scattered], not MSO[ordinal]

Remark 2. One aspect of Theorem 2 is that MSO[finite,cut] and MSO[ordinal] are equivalent. If we apply this fact to the domain ω, then cuts can be eliminated easily, and MSO[finite,cut] coincide with WMSO. Still over ω, MSO[ordinal] obviously coincide with MSO. Hence Theorem 2 implies that WMSO and MSO coincide over ω (in fact, the same argument is valid over any well ordered countable word). This non-trivial fact is usually established using the deep result of determinization of McNaughton [9] (other proofs involve weak alternating automata or algebra).

Theorem 3. *There are languages separating all situations not covered by Theorem 2.*

Proof (sketch). In fact, two among the five separating languages were given in Example 4: $L_{\mathrm{Ord}} \in$FO[cut]\WMSO and $L_{\mathrm{Scat}} \in$MSO[scattered]\MSO[ordinal].

WMSO\FO[cut]$\neq \emptyset$: The witnessing language is "the domain is of even finite length". It is the classical example of non-aperiodicity over finite words, and it works as well in this case.

MSO[ordinal]\(FO[cut]$\cup$WMSO) $\neq \emptyset$: For this, it is sufficient to take the disjoint union (for instance using disjoint alphabets) of a language in WMSO\FO[cut] and a language in FO[cut]\WMSO.

MSO\MSO[scattered]$\neq \emptyset$: Call a set X *perfectly dense* if all elements $x < y < z$ with $y \in X$ are such that (x, y) and (y, z) both intersect X. Said differently, all elements in X are limits from the left of elements from X, and symmetrically from the right. The language "there exists a set X of a-labelled positions which is perfectly dense" is obviously definable in MSO. Computing its syntactic $\circ$-monoid would yield four elements $1, a, b, 0$ with derived operations defined by $a \cdot a = a^{\omega} = a^{\omega *} = b \cdot b = b \cdot a = a \cdot b = b^{\omega} = b^{\omega *} = \{b\}^{\eta} = b$ and $\{a\}^{\eta} = \{a, b\}^{\eta} = 0$. The morphism sends a to a and b to b, and the accepting set is $\{0\}$. However, this language is not definable in MSO[scattered]: b is a shuffle idempotent which is not shuffle simple since $\{b\}^{\eta} = b = b \cdot a \cdot b$ and $\{a, b\}^{\eta} \neq b$. □

4 From logics to ∘-monoids

In this section, we show some of the results of the form "if a language $L \subseteq A^{\circ}$ is definable in logic $\mathcal{L}$, then its syntactic $\circ$-monoid satisfies property P" for suitable choices of $\mathcal{L}$ and P. The standard approach for such results is to use the technique of Ehrenfeucht-Fraïssé games. We adopt a different presentation here, making use of our fine understanding of $\circ$-monoids.

Let us first recall that all the logics we work with differ by their use of restricted set quantifiers. These restricted quantifiers are parameterized by a language $V \subseteq \{\in, \notin\}^{\circ}$. The quantifier $\exists^{V} X$ signifies "there exists a set of positions X which, when written as a labelling of the linear ordering yields a word in V". We have seen the language L_{Sing}, L_{Finite}, L_{Ord}, L_{Scat} that correspond to the quantifiers over singletons, finite sets, well ordered sets, and scattered sets.

Thus, the core step in each of these proofs consists in showing that the operation of restricted set quantifier preserves the property we are interested in when done at the level of $\circ$-monoids. Essentially, this looks as follows: "assume that L_{ϕ} is recognized by a $\circ$-monoid that has property P' then $L_{\exists^{V} X \phi}$ also has property P". Thus, we start by describing how $\exists^{V}$ behaves.

Let us just mention here that the existential quantifier is the crux of the problem, and that the other constructions involved (atomic predicates and boolean connectives) have also to be treated, but do not involve interesting arguments. We also have to verify the closure of the properties we are interested in under quotient of $\circ$-monoids. This last step is usually not necessary, but, since we did not choose to present the properties as identities, it has to be done explicitly.

4.1 Restricted quantifiers over ∘-monoids

Let us first recall how the existential set quantifier is implemented, from a language and algebraic theoretic point of view, and then refine this for restricted set quantifier.

Consider a language $L \in (A \times \{\in, \notin\})^\circ$. A word over this alphabet can be seen as a usual word over the alphabet A, enriched with the characteristic map of some set X: if a position belongs to X, then the second component is $\in$, otherwise it is $\notin$. The operation equivalent to existential set quantifier over such languages is $\mathit{Proj}(L)$ defined as:

$$\mathit{Proj}(L) = \big\{u_{|1} \in A^\circ \mid u \in L\big\} ,$$

where $u_{|1}$ denotes the word obtained by projecting each letter of u to its first component (similarly for $u_{|2}$). If furthermore L is recognized by some $\mathbf{M} = ((M, \pi), h, F)$, we define the new ∘-monoid $\mathcal{P}(\mathbf{M})$ to be $(\mathcal{P}(M), \pi)$, where

$$\text{for all } U \in (\mathcal{P}(M))^\circ, \quad \pi(U) = \{\pi(u) \mid u \in U\} ,$$

in which $u \in U$ holds if $\mathit{dom}(u) = \mathit{dom}(U)$ and for all $i \in \mathit{dom}(u)$, $u(i) \in U(i)$.

This construction is known to (1) produce a valid ∘-monoid, and (2) be such that $(\mathcal{P}(\mathbf{M}), h', F')$ recognizes $\mathit{Proj}(L)$ for $h'(a) = \{h(a, \in), h(a, \notin)\}$ and $F' = \{X \subseteq M \mid X \cap F \neq \emptyset\}$.

We present now a refinement of this construction, which furthermore restricts the range of the projection. Given a language $V \subseteq \{\in, \notin\}^\circ$ that represents the range of a restricted set quantifier, we define the *restricted projection* of L as:

$$\mathit{Proj}^V(L) = \big\{u_{|1} \in A^\circ \mid \text{for some } u \in L \text{ such that } u_{|2} \in V\big\} .$$

This operation is the language theoretic counterpart to the logical restricted quantifier $\exists^V$. Let us assume furthermore that V is recognized by some $(\mathbf{V}, g, E)$. We assume (and this will always be the case) that $\mathbf{V}$ has a zero 0, and that $0 \notin E$. We define the new ∘-monoid $\mathcal{P}_{\mathbf{V}}(\mathbf{M})$ to be (N, π), where

$$\begin{aligned} &\text{for all } U \in (\mathcal{P}(M \times V))^\circ, \quad \pi(U) = \{(\pi(u_{|1}), \pi(u_{|2})) \mid u \in U\} \setminus (M \times \{0\}) , \\ &\text{and} \quad N = \{\pi(U) \mid U \in \{\{(h(a, \in), g(\in)), (h(a, \notin), g(\notin))\} \mid a \in A\}^\circ\} . \end{aligned}$$

We can recognize in this construction the above powerset construction, applied to the ∘-monoid $\mathbf{M} \times \mathbf{V}$, from which all occurrences of the zero of $\mathbf{V}$ are removed as well all all non-reachable elements.

Lemma 2. *$\mathcal{P}_{\mathbf{V}}(\mathbf{M})$ is a ∘-monoid.*
If L is recognized by $(\mathbf{M}, h, F)$, then $\mathit{Proj}^V(L)$ is recognized by $(\mathcal{P}_{\mathbf{V}}(\mathbf{M}), h', F')$ where $h'(a) = \{(h(a, \in), g(\in)), (h(a, \notin), g(\notin))\}$ and $F' = \{A \mid A \cap (F \times E) \neq \emptyset\}$.

4.2 Establishing invariants

The core result in the translation from logics to ∘-monoids is the following.

Lemma 3. *Let* $\mathbf{M}$ *be a* $\circ$*-monoid.*

1. *If* $\mathbf{M}$ *satisfies* $\texttt{i}{\to}\texttt{gi}$ *then* $\mathcal{P}_{\mathbf{Sing}}(\mathbf{M})$ *satisfies* $\texttt{i}{\to}\texttt{gi}$.
2. *If* $\mathbf{M}$ *satisfies aperiodic then* $\mathcal{P}_{\mathbf{Cut}}(\mathbf{M})$ *satisfies aperiodic*[2].
3. *If* $\mathbf{M}$ *satisfies* $\texttt{oi}{\to}\texttt{gi}$ *then* $\mathcal{P}_{\mathbf{Fin}}(\mathbf{M})$ *satisfies* $\texttt{oi}{\to}\texttt{gi}$ *(resp.* $\texttt{o}^*\texttt{i}{\to}\texttt{gi}$*).*
4. *If* $\mathbf{M}$ *satisfies* $\texttt{sc}{\to}\texttt{sh}$ *then* $\mathcal{P}_{\mathbf{Ord}}(\mathbf{M})$ *satisfies* $\texttt{sc}{\to}\texttt{sh}$.
5. *If* $\mathbf{M}$ *satisfies* $\texttt{sh}{\to}\texttt{ss}$ *then* $\mathcal{P}_{\mathbf{Scat}}(\mathbf{M})$ *satisfies* $\texttt{sh}{\to}\texttt{ss}$.

Let us give some ideas about its proof. Let $\mathbf{N}$ be $\mathcal{P}_{\mathbf{V}}(\mathbf{M})$ where $\mathbf{V}$ is one of **Sing**, **Fin**, **Ord** or **Scat** (unfortunately, **Cut** having a different structure, it has to be treated separately).

Lemma 4. *There exists a* $\circ$*-monoid morphism* ρ *from* $\mathbf{N}$ *to* $\mathbf{M}$ *such that for all* $A \in N$*,* $(x, 1) \in A$ *if and only if* $x = \rho(A)$*.*

Proof. Essentially, the point is to prove that for all $A \in N$, there is one and only one $\rho(A)$ such that $(\rho(A), 1) \in A$. The fact that this ρ is a $\circ$-monoid morphism is then straightforward. For proving it, it is sufficient to do it for the neutral element $\{(1, 1)\}$, the image of each letter 'a' which happens to be $\{(h(a), 1),(h(a), s)\}$, and then show the preservation of the property under $\cdot, \omega, \omega^*$ and η. □

Let us show the simplest case of Lemma 3, the one for $\mathcal{P}_{\mathbf{Sing}}(\mathbf{M})$:

Lemma 5. *If a* $\circ$*-monoid* $\mathbf{M}$ *satisfies* $\texttt{i}{\to}\texttt{gi}$ *then* $\mathcal{P}_{\mathbf{Sing}}(\mathbf{M})$ *also does.*

Proof. Let E be an idempotent in $\mathbf{N} = \mathcal{P}_{\mathbf{Sing}}(\mathbf{M})$. Our goal is to show that it is gap insensitive.

Let $(x, y) \in E$. Since $E = E \cdot E$, there exists $(x_1, y_1),(x_2, y_2) \in E$ such that $x_1 \cdot x_2 = x$ and $y_1 \cdot y_2 = y$. Since $y \neq 0$, at least one among y_1, y_2 is equal to 1. Without loss of generality, let us assume it is y_1. In this case, according to Lemma 4, $x_1 = \rho(E)$. In particular, since ρ is a morphism, this means that x_1 is an idempotent. Thus we can use the assumption that $\mathbf{M}$ satisfies $\texttt{i}{\to}\texttt{gi}$ on it, and get that $x_1^\omega \cdot x_1^{\omega*} = x_1$. It follows that the word

$$\overbrace{(x_1,1)(x_1,1)\dots}^{\text{of domain } \omega}\ \overbrace{\dots(x_1,1)(x_1,1)}^{\text{of domain } \omega*}(x_2,y_2)$$

has also value (x, y) under π (componentwise), and as a consequence $(x, y) \in E^\omega \cdot E^{\omega*}$. We have proved $E \subseteq E^\omega \cdot E^{\omega*}$.

Conversely, consider some $(x, y) \in E^\omega \cdot E^{\omega*}$. This means that there exists a word u of the form

$$\overbrace{(x_1,y_1)(x_2,y_2)\dots}^{\text{of domain } \omega}\ \overbrace{\dots(x'_2,y'_2)(x'_1,y'_1)}^{\text{of domain } \omega*}$$

which evaluates (componentwise) to (x, y), with (x_i, y_i) and $(x'_i, y'_i) \in E$ for all $i \in \mathbb{N}$. If all $y = 1$, then its clear. Otherwise, there is at most one among the y_i's

[2] **Cut** is a $\circ$-monoid recognizing "cuts" that we omitted here for space reasons.

and the y_i''s which is not equal to 1. Without loss of generality (by symmetry), we can assume that it is y_j. According to Lemma 4, $x_i = \rho(E)$ for all $i \neq j$ and $x_i' = \rho(E)$ for all i. Since ρ is a morphism, $\rho(E)$ is also an idempotent. Thus we can use the assumption that $\mathbf{M}$ satisfies $\mathtt{i}{\rightarrow}\mathtt{gi}$. We obtain that $\rho(E)^\omega \cdot \rho(E)^{\omega *} = \rho(E)$. Thus, u evaluates to $(\rho(E), 1) \cdot (x_j, y_j) \cdot (\rho(E), 1) \in E^3 = E$. Hence $E^\omega \cdot E^{\omega *} \subseteq E$.

This terminates the proof that $\mathbf{N}$ satisfies $\mathtt{i}{\rightarrow}\mathtt{gi}$. □

5 Conclusion

In this paper we have characterized algebraically and effectively several natural sublogics of MSO. Unfortunately the most involved arguments, namely the translation from algebra to logic, were not addressed in this short abstract. These can be found in the appendix.

References

1. Alexis Bès and Olivier Carton. Algebraic characterization of FO for scattered linear orderings. In *CSL*, volume 12 of *LIPIcs*, pages 67–81. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2011.
2. Alexis Bès and Olivier Carton. Algebraic characterization of WMSO for scattered linear orderings. Personal communication, 2014.
3. J. Richard Büchi. On a decision method in restricted second order arithmetic. In *Logic, Methodology and Philosophy of Science (Proc. 1960 Internat. Congr .)*, pages 1–11. Stanford Univ. Press, Stanford, Calif., 1962.
4. Olivier Carton, Thomas Colcombet, and Gabriele Puppis. Regular languages of words over countable linear orderings. In *ICALP*, volume 6756, pages 125–136, 2011.
5. Rispal Chloé and Olivier Carton. An algebraic theory for regular languages of finite and infinite words. *Int. J. of Foundations of Comp. Sc.*, 16(04):767–786, 2005.
6. Thomas Colcombet. Monadic second-order logic and cuts in the backgrounds. In *CSR*, page Invited Paper, 2013.
7. Rosenstein J. G. *Linear Orderings.* Academic press, 1982.
8. Yuri Gurevich and Alexander Moshe Rabinovich. Definability and undefinability with real order at the background. *J. Symb. Log.*, 65(2):946–958, 2000.
9. Robert McNaughton. Testing and generating infinite sequences by a finite automaton. *Information and Control*, 9:521–530, 1966.
10. Robert McNaughton and Seymour Papert. *Counter-free Automata.* MIT Press, 1971.
11. Michael O. Rabin. Decidability of second-order theories and automata on infinite trees. *Trans. Amer. Math. Soc.*, 141:1–35, 1969.
12. Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.
13. Saharon Shelah. The monadic theory of order. *Ann. of Math. (2)*, 102(3):379–419, 1975.
14. Wolfgang Thomas. Languages, automata and logic. In Gregorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 3, chapter 7, pages 389–455. Springer, 1997.