Monadic Second-Order Logic with Arbitrary Monadic Predicates*

Nathanaël Fijalkow[1,2] and Charles Paperman[1]

[1] LIAFA, Paris 7
[2] University of Warsaw

Abstract. We study Monadic Second-Order Logic (**MSO**) over finite words, extended with (non-uniform arbitrary) monadic predicates. We show that it defines a class of languages that has algebraic, automata-theoretic and machine-independent characterizations. We consider the *regularity question*: given a language in this class, when is it regular? To answer this, we show a *substitution property* and the existence of a *syntactical predicate*.
We give three applications. The first two are to give simple proofs of the Straubing and Crane Beach Conjectures for monadic predicates, and the third is to show that it is decidable whether a language defined by an **MSO** formula with morphic predicates is regular.

1 Introduction

The Monadic Second-Order Logic (**MSO**) over finite words equipped with the linear ordering on positions is a well-studied and understood logic. It provides a mathematical framework for applications in many areas such as program verification, database and linguistics. In 1962, Büchi [5] proved the decidability of the satisfiability problem for **MSO** formulae.

Uniform Monadic Predicates. In 1966, Elgot and Rabin [9] considered extensions of **MSO** with uniform monadic predicates. For instance, the following formula

$$\forall x,\ \mathbf{a}(x) \iff x \text{ is prime},$$

describes the set of finite words such that the letters a appear exactly in prime positions. The predicate "x is a prime number" is a uniform monadic predicate on the positions, it can be seen as a subset of $\mathbb{N}$.

Elgot and Rabin were interested in the following question: for a uniform monadic predicate $\mathbf{P} \subseteq \mathbb{N}$, is the satisfiability problem of $\mathbf{MSO}[\leq, \mathbf{P}]$ decidable? A series of papers gave tighter conditions on $\mathbf{P}$, culminating to two final answers: in 1984, Semenov [19] gave a characterization of the predicates $\mathbf{P}$ such that

* The authors are supported by the project ANR 2010 BLAN 0202 02 (FREC). The first author is supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement 259454 (GALE) and 239850 (SOSNA). The second author is supported by Fondation CFM.

$\mathbf{MSO}[\leq, \mathbf{P}]$ is decidable, and in 2006, Rabinovich and Thomas [15, 17] proved that it is equivalent to the predicate $\mathbf{P}$ being effectively profinitely ultimately periodic.

Further questions on uniform monadic predicates have been investigated. For instance, Rabinovich [16] gave a solution to the Church synthesis problem for $\mathbf{MSO}[\leq, \mathbf{P}]$, for a large class of predicates $\mathbf{P}$.

In this paper, we consider non-uniform monadic predicates: such a predicate $\mathbf{P}$ is given, for each length $n \in \mathbb{N}$, by a predicate over the n first positions $\mathbf{P}_n \subseteq \{0, \dots, n-1\}$. The set $\mathcal{M}$ of these predicates contains the set $\mathcal{M}^{\text{unif}}$ of uniform monadic predicates.

Advice Regular Languages. We say that a language is *advice regular* if it is definable in $\mathbf{MSO}[\leq, \mathcal{M}]$. No computability assumptions are made on the monadic predicates, so this class contains undecidable languages.

Our first contribution is to give equivalent presentations of this class, which is a Boolean algebra extending the class of regular languages:

1. It has an equivalent automaton model: *automata with advice*.
2. It has an equivalent algebraic model: *one-scan programs*.
3. It has a machine-independent characterization, based on generalizations of Myhill-Nerode equivalence relations.

This extends the equivalence between automata with advice and Myhill-Nerode equivalence relations proved in [12] for the special case of uniform monadic predicates. We will rely on those characterizations to obtain several properties of the advice regular languages. Our main goal is the following regularity question: given an advice regular language L, when is L regular? To answer this question, we introduce two notions:

- The *substitution property*, which states that if a formula φ together with the predicate $\mathbf{P}$ defines a regular language $L_{\varphi,\mathbf{P}}$, then there exists a regular predicate $\mathbf{Q}$ such that $L_{\varphi,\mathbf{Q}} = L_{\varphi,\mathbf{P}}$.
- The *syntactical predicate* of a language L, which is the "simplest" predicate $\mathbf{P}_L$ such that $L \in \mathbf{MSO}[\leq, \mathbf{P}_L]$.

Our second contribution is to show that the class of advice regular languages has the substitution property, and that an advice regular language L is regular if and only if $\mathbf{P}_L$ is regular.

We apply these results to the case of morphic predicates [6], and obtain the following decidability result: given a language defined by an $\mathbf{MSO}$ formula with morphic predicates, one can decide whether it is regular.

Motivations from Circuit Complexity. Extending logics with predicates also appears in the context of circuit complexity. Indeed, a descriptive complexity theory initiated by Immermann [10] relates logics and circuits; it shows that a language is recognized by a Boolean circuit of constant depth and unlimited fan-in if and only if it can be described by a first-order formula with predicates (of any arity, so not only monadic ones), *i.e.* $\mathbf{AC^0} = \mathbf{FO}[\mathcal{N}]$.

This correspondence led to the study of two properties, which amount to characterize the regular languages (Straubing Conjecture) and the languages with a neutral letter (Crane Beach Conjecture) in several fragments of $\mathbf{FO}[\mathcal{N}]$. The Straubing Conjecture would, if true, give a deep understanding of many complexity classes inside $\mathbf{NC^1}$. Many cases of this conjecture are still open. On the other side, unfortunately the Crane Beach Conjecture does not hold in general, as shown by Barrington, Immermann, Lautemann, Schweikardt and Thérien [3]. On the positive side, both conjectures hold for uniform monadic predicates [3, 20].

Our third contribution is to give simple proofs of the both the Straubing and the Crane Beach Conjectures for monadic predicates relying on our previous characterizations.

Outline. The Section 2 gives characterizations of advice regular languages, in automata-theoretic, algebraic and machine-independent terms. In Section 3, we study the regularity question, and give two different answers: one through the substitution property, and the other through the existence of a syntactical predicate. The last section, Section 4, provides applications of our results: easy proofs that the Straubing and the Crane Beach Conjectures hold for monadic predicates and decidability of the regularity problem for morphic regular languages.

2 Advice Regular Languages

In this section, we introduce the class of advice regular languages and give several characterizations.

Predicates A monadic predicate $\mathbf{P}$ is given by $\mathbf{P} = (\mathbf{P}_n)_{n\in\mathbb{N}}$, where $\mathbf{P}_n \subseteq \{0,\ldots,n-1\}$. Since we mostly deal with monadic predicates, we often drop the word "monadic". In this definition the predicates are non-uniform: for each length n there is a predicate $\mathbf{P}_n$, and no assumption is made on the relation between $\mathbf{P}_n$ and $\mathbf{P}_{n'}$ for $n \neq n'$. A predicate $\mathbf{P}$ is uniform if there exists $\mathbf{Q} \subseteq \mathbb{N}$ such that for every n, $\mathbf{P}_n = \mathbf{Q} \cap \{0,\ldots,n-1\}$. We identify $\mathbf{P}$ and $\mathbf{Q}$, and see uniform predicates as subsets of $\mathbb{N}$.

For the sake of readability, we often define predicates as $\mathbf{P} = (\mathbf{P}_n)_{n\in\mathbb{N}}$ with $\mathbf{P}_n \subseteq \{0,1\}^n$. In such case we can see $\mathbf{P}$ as a language over $\{0,1\}$, which contains exactly one word for each length. Also, we often define predicates $\mathbf{P} = (\mathbf{P}_n)_{n\in\mathbb{N}}$ with $\mathbf{P}_n \in A^n$ for some finite alphabet A. This is not formally a predicate, but this amounts to define one predicate $\mathbf{P}^a$ for each letter a in A, with $\mathbf{P}^a_n(i) = 1$ if and only if $\mathbf{P}_n(i) = a$. This abuse of notations will prove very convenient. Similarly, any infinite word $w \in A^\omega$ can be seen as a uniform predicate.

Monadic Second-Order Logic The formulae we consider are monadic second-order (**MSO**) formulae, obtained from the following grammar:

$$\varphi \;=\; \mathbf{a}(x) \mid x \leq y \mid P(x) \mid \varphi \wedge \varphi \mid \neg\varphi \mid \exists x,\ \varphi \mid \exists X,\ \varphi$$

Here $x, y, z, \ldots$ are first-order variables, which will be interpreted by positions in the word, and $X, Y, Z, \ldots$ are monadic second-order variables, which will interpreted by sets of positions in the word. We say that $\mathbf{a}$ is a letter symbol, $\leq$ the ordering symbol and $P, Q, \ldots$ are the numerical monadic predicate symbols, often refered to as predicate symbols.

The notation $\varphi(P^1, \ldots, P^\ell, x^1, \ldots, x^n, X^1, \ldots, X^p)$ means that in φ, the predicate symbols are among $P^1, \ldots, P^\ell$, the free first-order variables are among $x^1, \ldots, x^n$ and the free second-order variables are among $X^1, \ldots, X^p$. A formula without free variables is called a sentence.

We use the notation $\overline{P}$ to abbreviate $P^1, \ldots, P^\ell$, and similarly for all objects (variables, predicate symbols, predicates).

We now define the semantics. The letter symbols and the ordering symbol are always interpreted in the same way, as expected. For the predicate symbols, the predicate symbol P is interpreted by a predicate $\mathbf{P}$. Note that P is a syntactic object, while $\mathbf{P}$ is a predicate used as the interpretation of P.

Consider $\varphi(\overline{P}, \overline{x}, \overline{X})$ a formula, u a finite word of length n, $\overline{\mathbf{P}}$ predicates interpreting the predicate symbols from $\overline{P}$, $\overline{\mathbf{x}}$ valuation of the free first-order variables and $\overline{\mathbf{X}}$ valuation of the free second-order variables. We define $u, \overline{\mathbf{P}}, \overline{\mathbf{x}}, \overline{\mathbf{X}} \models \varphi$ by induction as usual, with

$$u, \overline{\mathbf{P}}, \overline{\mathbf{x}}, \overline{\mathbf{X}} \models P(y) \quad \text{if} \quad \mathbf{y} \in \mathbf{P}_n \ .$$

A sentence $\varphi(\overline{P})$ and a tuple of predicates $\overline{\mathbf{P}}$ interpreting the predicate symbols from $\overline{P}$ define a language

$$L_{\varphi, \overline{\mathbf{P}}} = \{u \in A^* \mid u, \overline{\mathbf{P}} \models \varphi\} \ .$$

Such a language is called advice regular, and the class of advice regular languages is denoted by $\mathbf{MSO}[\leq, \mathcal{M}]$.

Automata with Advice We introduce automata with advice. Unlike classical automata, they have access to two more pieces of information about the word being read: its length and the current position. Both the transitions and the final states can depend on those two pieces of information. For this reason, they are (much) more expressive than classical automata, and recognize undecidable languages.

A non-deterministic automaton with advice is given by $\mathcal{A} = (Q, q_0, \delta, F)$ where Q is a finite set of states, $q_0 \in Q$ is the initial state, $\delta \subseteq \mathbb{N} \times \mathbb{N} \times Q \times A \times Q$ is the transition relation and $F \subseteq \mathbb{N} \times Q$ is the set of final states. In the deterministic case $\delta : \mathbb{N} \times \mathbb{N} \times Q \times A \to Q$.

A run over a finite word $u = u_0 \cdots u_{n-1} \in A^*$ is a finite word $\rho = q_0 \cdots q_n \in Q^*$ such that for all $i \in \{0, \ldots, n-1\}$, we have $(i, n, q_i, u_i, q_{i+1}) \in \delta$. It is accepting if $(n, q_n) \in F$.

One obtains a uniform model by removing one piece of information in the transition function: the length of the word. This automaton model is strictly weaker, and is (easily proved to be) equivalent to the one introduced in [12], where the automata read at the same time the input word and a fixed word called the advice. However, our definition will be better suited for some technical aspects: for instance, the number of Myhill-Nerode equivalence classes exactly correspond to the number of states in a minimal deterministic automaton.

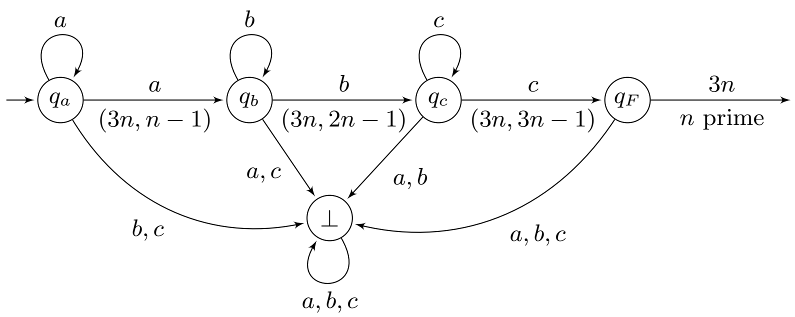

Fig. 1. The automaton for Example 2.1.

Example 2.1. The language $\{a^n b^n c^n \mid n \text{ is a prime number}\}$ is recognized by a (deterministic) automaton with advice. The automaton is represented in figure 1. It has five states, q_a, q_b, q_c, q_F and $\perp$. The initial state is q_a. The transition function is defined as follows:

$$
\begin{array}{lll}
\delta(i, 3n, q_a, a) & = q_a & \text{if } i < n-1 \\
\delta(n-1, 3n, q_a, a) & = q_b & \\
\delta(i, 3n, q_b, b) & = q_b & \text{if } n \leq i < 2n-1 \\
\delta(2n-1, 3n, q_b, c) & = q_c & \\
\delta(i, 3n, q_c, c) & = q_c & \text{if } 2n \leq i < 3n-1 \\
\delta(3n-1, 3n, q_c, c) & = q_F &
\end{array}
$$

All other transitions lead to $\perp$, the sink rejecting state. The set of final states is $F = \{(3n, q_F) \mid n \text{ is a prime number}\}$.

We mention another example, that appeared in the context of automatic structures [13]. They show that the structure $(\mathbb{Q}, +)$ is automatic with advice, which amounts to show that the language $\{\widehat{x} \sharp \widehat{y} \sharp \widehat{z} \mid z = x + y\}$, where $\widehat{x}$ denotes the factorial representation of the rational x, is advice regular.

One-scan Programs Programs over monoids were introduced in the context of circuit complexity [1]: Barrington showed that any language in $\mathbf{NC^1}$ can be computed by a program of polynomial length over a non-solvable group. Here we present a simplification introduced in [20], adapted to the context of monadic predicates.

A one-scan program is given by $P = (M, (f_{i,n} : A \to M)_{i,n\in\mathbb{N}}, S)$ where M is a finite monoid and $S \subseteq M$. The function $f_{i,n}$ is used to compute the effect of the i^{th} letter of an input word of length n. The program P accepts $u = u_0 \cdots u_{n-1}$ if $f_{0,n}(u_0) \cdots f_{n-1,n}(u_{n-1}) \in S$.

Note that this echoes the classical definition of recognition by monoids, where a morphism $f : A \to M$ into a finite monoid M recognizes the word $u = u_0 \cdots u_{n-1}$ if $f(u_0) \cdots f(u_{n-1}) \in S$. Here, a one-scan program uses different functions $f_{i,n}$, depending on the position i and the length of the word n.

Myhill-Nerode Equivalence Relations Let $L \subseteq A^*$ and $p \in \mathbb{N}$, we define two equivalence relations:

- $u \sim_L v$ if for all $w \in A^*$, we have $uw \in L \iff vw \in L$,
- $u \sim_{L,p} v$ if for all $w \in A^p$, we have $uw \in L \iff vw \in L$.

The relation $\sim_L$ is called the (classical) Myhill-Nerode equivalence relation. Recall that $\sim_L$ contains finitely many equivalence classes if and only if L is regular, *i.e.* $L \in \mathbf{MSO}[\leq]$.

Theorem 2.2 (Advice Regular Languages). *Let L be a language of finite words, the following properties are equivalent:*

(1) $L \in \mathbf{MSO}[\leq, \mathcal{M}]$,
(2) L is recognized by a non-deterministic automaton with advice,
(3) L is recognized by a deterministic automaton with advice,
(4) There exists $K \in \mathbb{N}$ such that for all $i, p \in \mathbb{N}$, the restriction of $\sim_{L,p}$ to words of length i contains at most K equivalence classes.
(5) L is recognized by a one-scan program,

In such case, we say that L is advice regular.

This extends the Myhill-Nerode theorem proposed in [12], which proves the equivalence between (3) and (4) for uniform predicates.

3 The Regularity Question

In this section, we address the following question: given an advice regular language, when is it regular? We answer this question in two different ways: first by showing a substitution property, and second by proving the existence of a syntactical predicate.

Note that the regularity question is not a decision problem, as advice regular languages are not finitely presentable, so we can only provide (non-effective) characterizations of regular languages inside the advice regular languages.

In the next section, we will show how these two notions answer the regularity question: first by proving that the Straubing property holds in this case, and second by proving the decidability of the regularity problem for morphic regular languages.

3.1 A Substitution Property

In this subsection, we prove a substitution property for $\mathbf{MSO}[\leq, \mathcal{M}]$.

We say that a predicate $\mathbf{P} = (\mathbf{P}_n)_{n\in\mathbb{N}}$ is regular if the language $\mathbf{P} \subseteq \{0,1\}^*$ is regular, defining the class $\mathcal{R}eg_1$ of regular *monadic* predicates (as defined in [14] and in [20]).

Theorem 3.1. *For all sentences $\varphi(\overline{P})$ in $\mathbf{MSO}[\leq, \mathcal{M}]$ and predicates $\overline{\mathbf{P}} \in \mathcal{M}$ such that $L_{\varphi,\overline{\mathbf{P}}}$ is regular, there exist $\overline{\mathbf{Q}} \in \mathcal{R}eg_1$ such that $L_{\varphi,\overline{\mathbf{Q}}} = L_{\varphi,\overline{\mathbf{P}}}$.*

The main idea of the proof is that among all predicates $\overline{\mathbf{Q}}$ such that $L_{\varphi,\overline{\mathbf{P}}} = L_{\varphi,\overline{\mathbf{Q}}}$, there is a minimal one with respect to a lexicographic ordering, which can be defined by an **MSO** formula. The key technical point is given by the following lemma, which can be understood as a regular choice function.

Lemma 3.2 (Regular Choice Lemma). *Let M be a regular language such that for all $k \in \mathbb{N}$, there exists a word $w \in M$ of length k. Then there exists $M' \subseteq M$ a regular language such that for all $k \in \mathbb{N}$, there exists exactly one word $w \in M'$ of length k.*

3.2 The Syntactical Predicate

In this subsection, we define the notion of syntactical predicate for an advice regular language. The word "syntactical" here should be understood in the following sense: the syntactical predicate $\mathbf{P}_L$ of L is the most regular predicate that describes the language L. In particular, we will prove that L is regular if and only if $\mathbf{P}_L$ is regular.

Let L be an advice regular language. We define the predicate $\mathbf{P}_L = (\mathbf{P}_{L,n})_{n\in\mathbb{N}}$. Thanks to Theorem 2.2, there exists $K \in \mathbb{N}$ such that for all $i, p \in \mathbb{N}$, the restriction of $\sim_{L,p}$ to words of length i contains at most K equivalence classes. Denote $Q = \{1, \ldots, K\}$ and $\Sigma = (Q \times A \to Q) \uplus Q$, where $Q \times A \to Q$ is the set of (partial) functions from $Q \times A$ to Q. We define $\mathbf{P}_{L,n} \in \Sigma^n$.

Let $i, n \in \mathbb{N}$. Among all words of length i, we denote by $u_1^{i,n}, u_2^{i,n}, \ldots$ the lexicographically minimal representants of the equivalence classes of $\sim_{L,n-i}$, enumerated in the lexicographic order:

$$u_1^{i,n} <_{\text{lex}} u_2^{i,n} <_{\text{lex}} u_3^{i,n} <_{\text{lex}} \cdots \tag{1}$$

In other words, $u_\ell^{i,n}$ is minimal with respect to the lexicographic order $<_{\text{lex}}$ among all words of length i in its equivalence class for $\sim_{L,n-i}$. Thanks to Theorem 2.2, there are at most K such words for each $i, n \in \mathbb{N}$.

We define $\mathbf{P}_{L,n}(i)$ (the i^{th} letter of $\mathbf{P}_{L,n}$) by:

$$\mathbf{P}_{L,n}(i)(\ell, a) = k \ \text{ if } \ u_\ell^{i,n} \cdot a \sim_{L,n-i-1} u_k^{i+1,n} \ , \text{ for } i < n \tag{2}$$

$$\mathbf{P}_{L,n}(n-1)(\ell) \ \text{ if } \ u_\ell^{n,n} \in L \ . \tag{3}$$

Intuitively, the predicate $\mathbf{P}_L$ describes the transition function with respect to the equivalence relations $\sim_{L,p}$. We now give an example.

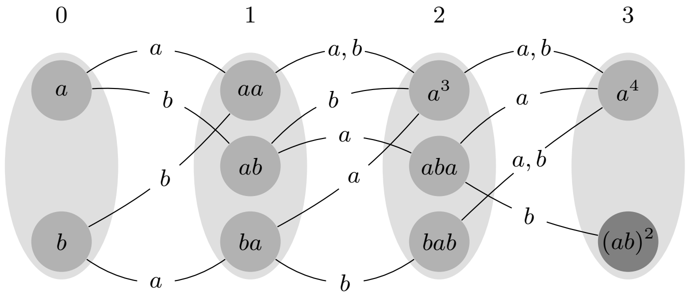

Fig. 2. The predicate $\mathbf{P}_L$ (here $\mathbf{P}_{L,4}$) for $L = (ab)^* + (ba)^*b$.

Example 3.3. Consider the language $L = (ab)^* + (ba)^*b$. We represent $\mathbf{P}_{L,4}$ in figure 2. Each circle represents an equivalence class with respect to $\sim_{L,4}$, inside words of a given length. For instance, there are three equivalence classes for words of length 3: a^3, aba and bab. Note that these three words are the minimal representants of their equivalence classes with respect to the lexicographic order. For the last position (here 3), the equivalence class of $(ab)^2$ (which is actually reduced to $(ab)^2$ itself) is darker since it belongs to the language L.

Theorem 3.4. *Let L be an advice regular language. Then L is regular if and only if $\mathbf{P}_L$ is regular.*

The proof is split in two lemmas, giving each direction.

Lemma 3.5. *Let L be an advice regular language. Then $L \in \mathbf{MSO}[\leq, \mathbf{P}_L]$.*

Lemma 3.6. *Let L be an advice regular language defined with the predicates $\overline{\mathbf{P}}$. Then $\mathbf{P}_L \in \mathbf{MSO}[\leq, \overline{\mathbf{P}}]$.*

4 Applications

In this section we show several consequences of Theorem 2.2 (characterization of the advice regular languages), Theorem 3.1 (a substitution property for advice regular languages) and Theorem 3.4 (a syntactical predicate for advice regular predicates).

The first two applications are about two conjectures, the Straubing and the Crane Beach Conjectures, introduced in the context of circuit complexity. We first explain the motivations for these two conjectures, and show simple proofs of both in the special case of monadic predicates.

The third application shows that one can determine, given an **MSO** formula with morphic predicates, whether it defines a regular language.

4.1 The Straubing and Crane Beach Conjectures

We first quickly define some circuit complexity classes. The most important here is $\mathbf{AC^0}$, the class of languages defined by boolean circuits of bounded depth and polynomial size. From $\mathbf{AC^0}$, adding the modular gates gives rise to $\mathbf{ACC}$. Finally, the class of languages defined by boolean circuits of logarithmic depth, polynomial size and fan-in 2 is denoted by $\mathbf{NC^1}$. Separating $\mathbf{ACC}$ from $\mathbf{NC^1}$ remains a long-standing open problem.

One approach to better understand these classes is through descriptive complexity theory, giving a perfect correspondence between circuit complexity classes and logical formalisms. Unlike what we did so far, the logical formalisms involved here use predicates of any arity (we focused on predicates of arity one). A k-ary predicate $\mathbf{P}$ is given by $(\mathbf{P}_n)_{n\in\mathbb{N}}$, where $\mathbf{P}_n \subseteq \{0,\dots,n-1\}^k$. We denote by $\mathcal{N}$ the class of all predicates, and by $\mathcal{R}eg$ the class of regular predicates as defined in [20].

Theorem 4.1 ([1, 4, 11]).

(1) $\mathbf{AC^0} = \mathbf{FO}[\mathcal{N}]$,
(2) $\mathbf{ACC} = (\mathbf{FO}+\mathbf{MOD})[\mathcal{N}]$.

Two conjectures have been formulated on the logical side, which aim at clarifying the relations between different circuit complexity classes. They have been stated and studied in special cases, we extrapolate them here to all fragments. Here the fragment $\mathbf{F}[\mathcal{P}]$ is described by a class $\mathbf{F}$ of formulae and a class $\mathcal{P}$ of predicates.

The first property, called the Straubing property, characterizes the regular languages (denoted by $\mathbf{REG}$) inside a larger fragment.

Definition 4.2 (Straubing Property). $\mathbf{F}[\mathcal{P}]$ *has the Straubing property if: all regular languages definable in* $\mathbf{F}[\mathcal{P}]$ *are also definable in* $\mathbf{F}[\mathcal{P}\cap\mathcal{R}eg]$. *In equation,* $\mathbf{F}[\mathcal{P}] \cap \mathbf{REG} = \mathbf{F}[\mathcal{P}\cap\mathcal{R}eg]$.

This statement appears for the first time in [2], where it is proved that $\mathbf{FO}[\mathcal{N}]$ has the Straubing property, relying on lower bounds for $\mathbf{AC^0}$ and an algebraic characterisation of $\mathbf{FO}[\mathcal{R}eg]$. Following this result, Straubing conjectures in [20] that $(\mathbf{FO}+\mathbf{MOD})[\mathcal{N}]$ and $\mathcal{B}\mathbf{\Sigma}_k[\mathcal{N}]$ have the Straubing property for $k \geq 1$. It has been shown that several fragments have the Straubing property, as for instance, $\mathbf{\Sigma_1}[\mathcal{N}]$, $\mathbf{FO}[\leq,\mathcal{M}^{\mathrm{unif}}]$ and $(\mathbf{FO}+\mathbf{MOD})[\leq,\mathcal{M}^{\mathrm{unif}}]$ (in [20]). We extend this result here, as a straightforward corollary of Theorem 3.1.

Theorem 4.3. *All fragments* $\mathbf{F}[\leq,\mathcal{M}]$ *have the Straubing property.*

In particular, for all $k \geq 1$, $\mathcal{B}\mathbf{\Sigma}_k[\leq,\mathcal{M}]$ has the Straubing property. This result is, to the best of our knowledge, the first intermediary result towards a proof of the Straubing Conjecture for $\mathcal{B}\mathbf{\Sigma}_k[\mathcal{N}]$.

The second property, called the Crane Beach property, characterizes the languages having a neutral letter, and has been proposed by Thérien for the special case of first-order logic. We say that a language L has a neutral letter e if for all words u, v, we have $uv \in L$ if and only if $uev \in L$.

Definition 4.4 (Crane Beach Property). $\mathbf{F}[\mathcal{P}]$ *has the Crane Beach property if: all languages having a neutral letter definable in* $\mathbf{F}[\mathcal{P}]$ *are definable in* $\mathbf{F}[\leq]$.

Unfortunately, the Crane Beach property does not hold in general.

Theorem 4.5 ([3, 18]). *There exists a non-regular language having a neutral letter definable in* $\mathbf{FO}[\mathcal{N}]$.

A deeper understanding of the Crane Beach property specialized to first-order logic can be found in [3]. In particular, it has been shown that $\mathbf{FO}[\leq, \mathcal{M}^{\mathrm{unif}}]$ has the Crane Beach property. Here we obtain the following result as a simple corollary of Theorem 2.2.

Theorem 4.6. $\mathbf{MSO}[\leq, \mathcal{M}]$ *has the Crane Beach property.*

4.2 Morphic Regular Languages

In this subsection, we apply Theorem 3.4 to the case of morphic predicates, and obtain the following result: given an **MSO** formula with morphic predicates, it is decidable whether it defines a regular language.

The class of morphic predicates was first introduced by Thue in the context of combinatorics on words, giving rise to the HD0L systems. Formally, let A, B be two finite alphabets, $\sigma : A^* \to A^*$ a morphism, $a \in A$ a letter such that $\sigma(a) = a \cdot u$ for some $u \in A^+$ and $\varphi : A^* \to B^*$ a morphism. This defines the sequence of words $\varphi(a), \varphi(\sigma(a)), \varphi(\sigma^2(a)), \ldots$, which converges to a finite or infinite word. An infinite word obtained in this way is said morphic.

We see morphic words as predicates, and denote by HD0L the class of morphic predicates. The languages definable in $\mathbf{MSO}[\leq, \mathrm{HD0L}]$ are called morphic regular.

Theorem 4.7. *The following problem is decidable: given L a morphic regular language, is L regular?*

The proof of this theorem goes in two steps: first, we reduce the regularity problem for a morphic regular language L to deciding the ultimate periodicity of $\mathbf{P}_L$, and second, we show that $\mathbf{P}_L$ is morphic. Hence we rely on the following result: given a morphic word, it is decidable whether it is ultimately periodic. The decidability of this problem was conjectured 30 years ago and proved recently and simultaneously by Durand and Mitrofanov [8].

The first step is a direct application of Theorem 3.4. For the second step, observe that thanks to Lemma 3.6, we have $\mathbf{P}_L \in \mathbf{MSO}[\leq, \mathrm{HD0L}]$. We conclude with the following result from [7].

Lemma 4.8. HD0L *is closed under* **MSO***-interpretations, i.e. if* $\mathbf{P}$ *is an infinite word such that* $\mathbf{P} \in \mathbf{MSO}[\leq, \mathrm{HD0L}]$*, then* $\mathbf{P} \in \mathrm{HD0L}$.

Acknowledgments

This paper and its authors owe a lot to the numerous fruitful discussions we had with Olivier Carton, Thomas Colcombet, Sam van Gool and Jean-Éric Pin. The same goes for the anonymous referees, who improved the paper by their constructive comments.

References

1. David A. Mix Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *Journal of Computer and System Sciences*, 38(1):150–164, 1989.
2. David A. Mix Barrington, Kevin Compton, Howard Straubing, and Denis Thérien. Regular languages in NC^1. *Journal of Computer and System Sciences*, 44(3):478–499, 1992.
3. David A. Mix Barrington, Neil Immerman, Clemens Lautemann, Nicole Schweikardt, and Denis Thérien. First-order expressibility of languages with neutral letters or: The Crane Beach conjecture. *Journal of Computer and System Sciences*, 70(2):101–127, 2005.
4. David A. Mix Barrington and Denis Thérien. Finite monoids and the fine structure of NC^1. *Journal of the Association for Computing Machinery*, 35(4):941–952, 1988.
5. J. Richard Büchi. On a decision method in restricted second-order arithmetic. In *Proceedings of the 1st International Congress of Logic, Methodology, and Philosophy of Science, CLMPS'60*, pages 1–11. Stanford University Press, 1962.
6. Olivier Carton and Wolfgang Thomas. The monadic theory of morphic infinite words and generalizations. *Information and Computation*, 176(1):51–65, 2002.
7. F. M. Dekking. Iteration of maps by an automaton. *Discrete Mathematics*, 126(1-3):81–86, 1994.
8. Fabien Durand. Decidability of the HD0L ultimate periodicity problem. *RAIRO Theor. Inform. Appl.*, 47(2):201–214, 2013.
9. Calvin C. Elgot and Michael O. Rabin. Decidability and undecidability of extensions of second (first) order theory of (generalized) successor. *Journal of Symbolic Logic*, 31(2):169–181, 1966.
10. Neil Immerman. Languages that capture complexity classes. *SIAM Journal of Computing*, 16(4):760–778, 1987.
11. Michal Koucký, Clemens Lautemann, Sebastian Poloczek, and Denis Thérien. Circuit Lower Bounds via Ehrenfeucht-Fraissé Games. In *IEEE Conference on Computational Complexity*, pages 190–201, 2006.
12. Alex Kruckman, Sasha Rubin, John Sheridan, and Ben Zax. A myhill-nerode theorem for automata with advice. In Marco Faella and Aniello Murano, editors, *GandALF*, volume 96 of *EPTCS*, pages 238–246, 2012.
13. André Nies. Describing groups. *Bulletin of Symbolic Logic*, 13:305–339, 9 2007.
14. Pierre Péladeau. Logically defined subsets of $\mathbb{N}^k$. *Theoretical Computer Science*, 93(2):169–183, 1992.
15. Alexander Rabinovich. On decidability of monadic logic of order over the naturals extended by monadic predicates. *Information and Computation*, 205(6):870–889, 2007.
16. Alexander Rabinovich. The Church problem for expansions of $(\mathbb{N}, <)$ by unary predicates. *Information and Computation*, 218:1–16, 2012.

17. Alexander Rabinovich and Wolfgang Thomas. Decidable theories of the ordering of natural numbers with unary predicates. In Zoltán Ésik, editor, *CSL*, volume 4207 of *Lecture Notes in Computer Science*, pages 562–574. Springer, 2006.
18. Nicole Schweikardt. *On the Expressive Power of First-Order Logic with Built-In Predicates*. PhD thesis, Gutenberg-Universtät in Mainz, 2001.
19. Alexei L. Semenov. Decidability of monadic theories. In Michal Chytil and Václav Koubek, editors, *MFCS*, volume 176 of *Lecture Notes in Computer Science*, pages 162–175. Springer, 1984.
20. Howard Straubing. *Finite automata, formal logic, and circuit complexity.* Birkhäuser Boston Inc., Boston, MA, 1994.