Adding modular predicates to first-order fragments

Luc Dartois, Charles Paperman

*Discrete Mathematics and Theoretical Computer Science*DMTCS vol. **(subm.)**:1, by the authors, #rev

Adding modular predicates to first-order fragments

Luc Dartois[1,2] Charles Paperman[3]

[1] *LIF, Aix-Marseille Université, France*
[2] *Université Libre de Bruxelles, Belgium*
[3] *Warsaw University, Poland*

We investigate the decidability of the definability problem for fragments of first order logic over finite words enriched with modular predicates. Our approach aims toward the most generic statements that we could achieve, which successfully covers the quantifier alternation hierarchy of first order logic and some of its fragments. We obtain that deciding this problem for each level of the alternation hierarchy of both first order logic and its two-variable fragment when equipped with all regular numerical predicates is not harder than deciding it for the corresponding level equipped with only the linear order and the successor. For two-variable fragments we also treat the case of the signature containing only the order and modular predicates.

Relying on some recent results, this proves the decidability for each level of the alternation hierarchy of the two-variable first order fragment while in the case of the first order logic the question remains open for levels greater than two.

The main ingredients of the proofs are syntactic transformations of first order formulas as well as the algebraic framework of finite categories.

Keywords: First order logic, automata theory, semigroup, modular predicates

1 Introduction

The equivalence between regular languages and automata (Rabin and Scott, 1959) as well as monadic second order logic (Büchi, 1960) and finite monoids (Nerode, 1958) was the start of a domain of research that is still active today. In this article, we are interested in the logic on finite words, and more precisely the question we address is the *definability problem* for fragments of logic. Fragments of logic are defined as sets of monadic second order formulas satisfying some restrictions, and are equipped with a set of predicates called a *signature*. Then the definability problem of a fragment of logic $\mathcal{F}$ consists in deciding if a regular language can be defined by a formula of $\mathcal{F}$.

This question has already been considered and solved in many cases where the signature contains only the predicate $<$, which denotes the linear order over the positions of the word. For instance, a celebrated result by Schützenberger (1965) and McNaughton and Papert (1971) gave an effective algebraic characterization of languages definable by first order formulas. The decidability has often been achieved through algebraic means, showing a deep connection between

algebraic and logical properties of a given regular language. This is the approach privileged in this article.

We investigate the question of the behaviour of the decidability of some fragments when their signature is enriched with *modular predicates*. These predicates allow to specify the congruence of the position of a variable modulo an integer. They form with the order and the local predicates the set of *regular numerical predicates*. These predicates are exactly the formulas of monadic second order logic without letter predicates. Intuitively they correspond to the maximal class of *numerical predicates* that can enrich the signature of a fragment of **MSO**, while keeping the definable languages regular. This question was already considered in the case of first order logic (**FO**) by Barrington et al. (1992) and one of its fragments, the formulas without quantifier alternation, by Péladeau (1992).

The enrichment by regular numerical predicates arose in the context of the *Straubing's conjectures* (Straubing, 1994). Roughly speaking, these conjectures state that deciding the definability of a regular language in a fragment of enriched logic corresponds to deciding its circuit complexity. It is known (Péladeau, 1992; Straubing, 1994) that an enrichment of the classical fragments by regular numerical predicates is equivalent to an enrichment by the signature $[<, +1, \mathrm{MOD}]$, where +1 denotes the *local predicates* and MOD the *modular predicates*. A first step toward the study of fragments of logic with these predicates was initiated by Straubing (1985). He obtained that adding the local predicates preserves the decidability for a large number of fragments. As a corollary of this work, Straubing obtained that the decidability of the alternation hierarchy of first order logic ($\mathcal{B}\mathbf{\Sigma}_k$) equipped with $[<, +1]$ reduces to the decidability of the simpler one $[<]$. More recently, Kufleitner and Lauser (2013) proved the decidability of the alternation hierarchy of the two-variable first order fragment ($\mathbf{FO}^2_k$) equipped with $[<, +1]$ by extending the recent results by Krebs and Straubing (2012) and Kufleitner and Weil (2012) on the decidability of this hierarchy with $[<]$.

In this context, the case of modular predicates is poorly understood. The study of this enrichment was first considered for first order logic by Barrington et al. (1992), and had been extended to the first level of its alternation hierarchy with the successor predicate by Péladeau (1992), and later without it by Chaubard et al. (2006). The enrichment by a finite set of modular predicate was considered by Ésik and Ito (2003). Finally, the authors provided a characterization of the two-variable first order logic over the signature $[<, \mathrm{MOD}]$ (Dartois and Paperman, 2013). In this paper, we focus on the enrichment by all regular predicates as well as the question of the enrichment by modular predicates only. This latter one surprisingly turns out to be more intricate.

To study this enrichment in a generic setting, we offer a definition of fragment as a set of formulas satisfying some syntactic properties. This allows for some generic proofs instead of a one by one situation. The main applications of our theorems are then the *quantifier alternations* hierarchies of the first order logic and its two-variable counterpart. Our main results state that for both of these hierarchies, the decidability of each level equipped with regular numerical predicates reduces to decidability of the same level with the signature $[<, +1]$. Then by using the recent decidability result of Kufleitner and Lauser (2013), as well as the decidability of $\mathcal{B}\mathbf{\Sigma}_2[<]$ by Place and Zeitoun (2014), we deduce that the fragments $\mathbf{FO}^2[<, \mathrm{MOD}]$ and $\mathbf{FO}^2_k[\mathrm{Reg}]$, for any positive k, as well as $\mathcal{B}\mathbf{\Sigma}_2[\mathrm{Reg}]$ are decidable. Our settings also reproves known results and apply to fragments of first order with small signatures.

Proofs methods. The proofs of the main results can be decomposed in two major steps. The first part is rather classical and shows that the information given by a finite number of modular predicates can be put into the alphabet and thus we can reduce the problem to a question on the fragment over a bigger alphabet. The second part is dedicated to finding a systematic way to select, for a given regular language and a fragment, a finite number of modular predicates that can serve as witnesses of its definability. This is done through the use of the algebraic framework of varieties, using two mains approaches. The first one uses finite categories and the global of a variety, while the second one introduces a new notion for varieties of semigroups that we call the infinitely testable property. Under some assumptions, we show that this property allows us to find such a witness set for modular predicates.

Organization of the paper. The next section is dedicated to the basic logical and algebraic definitions, and the main applications of our results to logic are presented in Section 4. Then Section 3 deals with adding a finite number of modular predicates. This is done through an easy reduction to adding predicates modulo a given congruence. In Section 5, we then deal with the delay problem, which can be quickly stated as computing a finite set of congruences that can serve as a witness for the definability problem of a language. More specifically, we first introduce the framework of categories as an extension of the monoids theory, and use it to prove a delay for differents classes of fragments. In Subsections 5.2 and 5.3, we rely on an algebraic description of the global of a variety, which is a variety of finite categories. Then Subsection 5.4 solves the delay for a class of fragments satisfying a given algebraic property, the so-called *infinitely testable property*.

2 Preliminary definitions

2.1 Languages and Logic

We consider the monadic second order logic on finite words $\mathbf{MSO}[<]$ as usual (see Straubing (1994) for example). We denote by A an *alphabet* and by a a *letter* of A. A word u over an alphabet A is a set of labelled positions ordered from 0 to $|u|-1$, where $|u|$ is an integer denoting the *length* of u. The set of words over A is denoted A^* and a subset L of A^* is called a *language*. We also denote by A^+ the set of non-empty words. A language is said to be *defined* by a formula if it corresponds exactly to the set of words that satisfy this formula. It is said to be *regular* if it is defined by a $\mathbf{MSO}[<]$ formula. When syntactic restrictions are applied to $\mathbf{MSO}[<]$, one defines fragments of logic that characterize subclasses of regular languages. The most well-known fragment is probably the first order logic, whose expressive power was characterized thanks to the results of McNaughton and Papert (1971) and Schützenberger (1965). The first order logic itself gave birth to its own zoo of fragments. These were defined using syntactical restrictions such as limiting the number of variables, or by enrichment of its signature. A fragment $\mathcal{F}$ with signature σ will be denoted $\mathcal{F}[\sigma]$ and will refer to the formulas as well as the class of languages it defines.

We first define the different signatures that will appear through this paper, and then formally define the quantifier alternation hierarchies, as they form the main focus of the applications of our theorems.

Signatures. We are interested in regular numerical predicates, which are numerical predicates that can only define regular languages. Simultaneously, Straubing (1994) and Péladeau (1992) defined three sets of regular numerical predicates that can be used as a base for all the regular numerical predicates. The first set is the singleton order $\{<\}$ which is a binary predicate corresponding to the natural order on the positions of the input word. The second set is $\{\mathbf{min}, \mathbf{max}, \mathbf{S_k}\}$ and is called the *local predicates*. The predicates **min** and **max** are unary predicates that are satisfied respectively on the first and last positions. The predicate $\mathbf{S_k}$, the k^{th}-*successor*, is a binary predicate satisfied if the second variable quantifies the k^{th}-successor of the first one.

Example 1. *The formula* $\exists x \exists y\ \mathbf{min}(x) \wedge \mathbf{S}(x,y) \wedge \mathbf{a}(x) \wedge \mathbf{a}(y)$ *defines the regular language* aaA^*.

We alternatively use the *descriptive local predicates*. These predicates are of the form $\mathbf{a}(x+k)$ (resp. $\mathbf{a}(\min + k)$, $\mathbf{a}(\max - k)$) for $k \geqslant 0$, holding if the position at $x+k$ (resp. $min+k$, $max-k$) is labelled by an a.

Example 2. *The previous formula can be rewrite by the following quantifier-free formula:* $\mathbf{a}(\min) \wedge \mathbf{a}(\min +1)$.

Most of the time, both *descriptive* and classical local predicates provides the same expressive power. However the descriptive predicates are proved to be more convenient for abstract fragments since they don't bound two variables together. For the sake of simplicity we will denote in the following by +1 this class of descriptive local predicates. This notation is justified thanks to the close relation between descriptive local predicates and the successor function. Also note that the presence or absence of the equality predicate is important since $\mathbf{FO}[+1]$ is strictly less expressive than $\mathbf{FO}[=,+1]$.

Finally, we define, for each positive integer d, the *modular predicates on* d, denoted MOD^d, as the set, for $i < d$, of predicates $\mathbf{MOD_i^d}(x)$ which are unary predicates satisfied if the position quantified by x is congruent to i modulo d, and the predicates $\mathbf{D_i^d}$ which are constants holding if the length of the input word is congruent to $i \bmod d$. We denote by MOD the union of the classes MOD^d, for any positive d.

Example 3. *The language* $(A^2)^* aA^*$ *is defined by the formula:* $\exists x\ \mathbf{a}(x) \wedge\ \mathbf{MOD_0^2}(x)$.

The signatures that we will consider for our fragments are unions of these three sets of regular numerical predicates, and will always contain the letter predicates. Abusing notations, we will also write $\text{Reg} = \{<\} \cup +1 \cup \text{MOD}$.

Fragments. A *fragment of logic* $\mathcal{F}[\sigma]$ with signature σ is a set of closed formulas of $\mathbf{MSO}[\sigma]$ that contains the quantifier-free formulas and that is closed under the following operations :

Conjunction If φ and ψ are formulas of $\mathcal{F}$, then $\varphi \wedge \psi$ is also a formula of $\mathcal{F}$.

Disjunction If φ and ψ are formulas of $\mathcal{F}$, then $\varphi \vee \psi$ is also a formula of $\mathcal{F}$.

Quantifier-free substitutions If φ is a formula of $\mathcal{F}$ and $\psi(x_1, \ldots, x_n)$ a quantifier-free subformula of φ with free variables $x_1, \ldots, x_n$, then any formula obtained by replacing $\psi(x_1, \ldots, x_n)$ by another quantifier-free formula with the same set of free variables is also in $\mathcal{F}$.

If $\mathcal{F}[\sigma]$ is a fragment of logic and σ' is a class of predicates, then the *enrichment* of $\mathcal{F}[\sigma]$ by σ' is denoted by $\mathcal{F}[\sigma, \sigma']$ and corresponds to the closure of $\mathcal{F}[\sigma]$ under the quantifier-free

substitutions, where predicates range over the signature $\sigma \cup \sigma'$. As a closed formula defines a language, a fragment of logic defines a class of languages. Abusing notations, we will denote by $\mathcal{F}[\sigma]$ a fragment of logic, as well as the class of languages it recognizes. It is worth noting that Kufleitner and Lauser (2012) defined another notion of fragment of logic as sets of formulas closed under some syntactical substitutions ensuring algebraic characterisation of the fragment.

The fragment $\mathbf{FO}^2$ is the subclass of formulas of $\mathbf{FO}$ using only two symbols of variables which can be reused (see Example 4). Here, the class of languages defined by $\mathbf{FO}^2[<]$ is strictly contained in $\mathbf{FO}^2[<, +1]$ and $\mathbf{FO}^2[<, \mathrm{MOD}]$ (see Thérien and Wilke (1999); Dartois and Paperman (2013)).

Example 4. *The language $A^*aA^*bA^*aA^*$ can be described by the first order formula*

$$\exists x \exists y \exists z \ x < y < z \wedge \mathbf{a}(x) \wedge \mathbf{b}(y) \wedge \mathbf{a}(z) \ .$$

This formula uses three variables x, y and z. However, by reusing x we get an equivalent formula that uses only two variables:

$$\exists x \ \mathbf{a}(x) \wedge \Big(\exists y \ x < y \wedge \mathbf{b}(y) \wedge \big(\exists x \ y < x \wedge \mathbf{a}(x) \big) \Big) \ . \tag{a}$$

Alternation hierarchies. Given a first order formula, one can compute a prenex normal form using the De Morgan's laws. We define the *quantifier alternation depth* of a formula as the number of blocks of quantifiers $\forall$ and $\exists$ in its prenex normal form. For example, the formula $\exists x \exists y \forall z \ x < z < y \wedge \mathbf{a}(x) \wedge \mathbf{a}(y) \wedge \mathbf{c}(z)$ has a quantifier depth of 2. It describes the language $A^*ac^*aA^*$. Then given a signature σ and a positive integer k, we denote by $\mathcal{B}\mathbf{\Sigma}_k[\sigma]$ the set of prenex normal formulas of $\mathbf{FO}[\sigma]$ whose quantifier depth is smaller or equal to k. They form the levels of the *quantifier alternation hierarchy* over $\mathbf{FO}[\sigma]$.

When σ is reduced to $\{<\}$, this hierarchy is called the Straubing-Thérien hierarchy (Straubing, 1981; Thérien, 1981). Only the first (Simon, 1975) and second (Place and Zeitoun, 2014) levels are known to be decidable. For $\sigma = \{<\} \cup +1$, this hierarchy is called the Dot-Depth hierarchy (Cohen and Brzozowski, 1971). The decidability of each level reduces to the decidability of the corresponding level of the Straubing-Thérien hierarchy (Straubing, 1985). In both cases, the hierarchies are known to be strict, and cover all Star-Free languages. In this article, we also consider the alternation hierarchy of $\mathbf{FO}^2$. To define formally the number of alternations of a formula, we cannot rely on the prenex normal form since the construction increases the number of variables. In particular, remark that $\mathbf{FO}^2[<]$ is equivalent to $\mathbf{\Sigma_2}[<] \cap \mathbf{\Pi}_2[<]$ which is a subclass of $\mathcal{B}\mathbf{\Sigma}_2[<]$ (Diekert et al., 2008). That said, the number of alternations is still a relevant parameter that could be defined as follows: Consider the parse tree naturally associated to a formula. For instance, (a) has $\exists$ as a root and the atomic formulas as the leaves. In a two-variable first order formula we count the maximal number of alternations appearing on a branch, i.e. between the root and a leaf, once the negations have been pushed on to the leaves. A more precise definition can be found in Weis and Immerman (2009). We denote by $\mathbf{FO}^2_k[\sigma]$ the formulas of $\mathbf{FO}^2[\sigma]$ that have at most $k-1$ quantifier alternations. The hierarchy induced by $\mathbf{FO}^2_k[<]$ is known to be strict (Weis and Immerman, 2009) and its definability problem is decidable (Krebs and Straubing, 2012; Kufleitner and Weil, 2012). Note that the hierarchy $\mathbf{FO}^2_k[<, +1]$ is also known to be decidable (Kufleitner and Lauser, 2013).

Remark: The classes of formulas $\mathbf{FO}$ and $\mathbf{FO}^2$ as well as each level of the alternation hierarchies are fragments of $\mathbf{MSO}$ as defined previously.

2.2 Varieties of languages, monoids and semigroups

We quickly present here the fundamental notions used by the article and refer the reader to the book of Pin (1997) for a detailed approach. A (finite) *semigroup* is a finite set equipped with an associative internal law. A semigroup with a neutral element for this law is called a *monoid*. Recall that a semigroup S *divides* another semigroup T if S is a quotient of a subsemigroup of T. This defines a partial order on finite semigroups. Given a finite semigroup S, an element e of S is *idempotent* if $ee = e$. We denote by $E(S)$ the set of idempotents of S. For any element x of S, there exists a positive integer n such that x^n is idempotent. We call this element the *idempotent power* of x and denote it by x^ω. One can check that the application $x \to x^\omega$ is well defined.

A semigroup S recognizes a language L over an alphabet A via a *morphism* $\eta : A^+ \to S$. Given a regular language L, we can compute its *syntactic semigroup* as the smallest semigroup that recognizes L, in the sense of division. A subset T of S is an *ideal* if the sets TS and ST are both included in T. A (pseudo-)*variety* of semigroups (resp. monoids) is a non empty class of finite semigroups (resp. monoids) closed under division and finite product. Finally, a local monoid of S is a monoid of the form eSe where e is an idempotent of S.

A fragment of logic is *characterized* by a variety if they recognize the same languages. By extension, a variety $\mathbf{V}$ will also refer to the class of languages it recognizes. The most famous example is the equality $\mathbf{FO}[<] = \mathbf{A}$ (McNaughton and Papert, 1971; Schützenberger, 1965), where $\mathbf{A}$ denotes the class of aperiodic semigroups, which are finite semigroups that are not divided by any group. As for $\mathbf{FO}[<]$, the definability problem for a fragment of logic has often been solved thanks to an algebraic characterization (Simon (1975); Thérien (1981); Thérien and Wilke (1999) for example). This decidability is sometimes obtained through *profinite equations*. We refer the reader to Pin (2009) for a survey on the profinite background. The algebraic characterisations of most the fragments that we consider are given in Figure 1.

Fragment	Variety	Equations
$\mathbf{FO}[<]$	$\mathbf{A}$	$x^\omega = x^{\omega+1}$
$\mathbf{FO}[=]$	$\mathbf{ACom}$	$x^\omega = x^{\omega+1}$, $xy = yx$
$\mathbf{FO}^2[<]$	$\mathbf{DA}$	$(xy)^\omega = (xy)^\omega x (xy)^\omega$
$\mathbf{FO}^1[\varnothing]$	$\mathbf{J}_1$	$x^2 = x$, $xy = yx$
$\mathcal{B}\mathbf{\Sigma}_1[<]$	$\mathbf{J}$	$y(xy)^\omega = (xy)^\omega = (xy)^\omega x$
$\mathbf{FO}^2_k[<]$	$\mathbf{V}_k$	See Example 6

Fig. 1: Algebraic characterisations

Stability index. One important tool to study modular predicates is the stability index. For a monoid morphism $\varphi : A^* \to M$, the set $\varphi(A)$ is an element of the powerset monoid of M. As such it has an idempotent power. The *stability index* of a morphism is the least positive integer s such that $\varphi(A^s) = \varphi(A^{2s})$. This set forms a subsemigroup called the *stable semigroup* of φ. The set $\varphi((A^s)^*)$ is called the *stable monoid* of φ. The stable monoid (resp. semigroup) of a regular

language is the stable monoid (resp. semigroup) of its syntactic morphism.

3 Adding finitely many modular predicates

We consider here the question of adding the modular predicates associated to a finite set of congruences. First, let us remark that if $\mathcal{F}[\sigma]$ is a fragment of logic and d and p are two positive integers, then $\mathcal{F}[\sigma, \mathrm{MOD}^d, \mathrm{MOD}^p] \subseteq \mathcal{F}[\sigma, \mathrm{MOD}^{dp}]$. This can be proved by some basic arithmetic reasoning and some quantifier-free substitutions. Then as a formula only uses a finite number of modular predicates, for any language of $\mathcal{F}[\sigma, \mathrm{MOD}]$, there exists an integer d such that it belongs to $\mathcal{F}[\sigma, \mathrm{MOD}^d]$. The consequence is that adding a finite set of modular predicates is equivalent to adding the predicates relating to one specific congruence. The remainder of this section deals with this question.

3.1 Alphabet enriched by modular counting

In order to deal with modular predicates, we now define enriched modular alphabets. These notions come naturally in the context of *wreath product* and instantiated for instance in Dartois and Paperman (2013, 2015). We now fix a positive integer d and an alphabet A. Let $\mathbb{Z}_d$ be the cyclic group of order d.

Definition 1 (Enriched alphabet)**.** *We call the set $A_d = A \times \mathbb{Z}_d$ the* enriched alphabet *of A, and we denote by $\pi_d : A_d^* \to A^*$ the projection defined by $\pi_d(a, i) = a$ for each $(a, i) \in A_d$. For example, the word $(a,2)(b,1)(b,2)(a,0)$ is an enriched word of abba for $d = 3$. We say that abba is the* underlying word *of $(a,2)(b,1)(b,2)(a,0)$.*

Definition 2 (Well-formed words)**.** *A word $(a_0, i_0)(a_1, i_1)\cdots(a_n, i_n)$ of A_d^* is* well-formed *if for $0 \leqslant j \leqslant n$, $i_j = j \bmod d$. We denote by K_d the set of all well-formed words of A_d^*. We also note $A_d(i, j)$ the set of well-formed factor such that the first letter is labelled by i and the last by j.*

For any $i < d$, let $\alpha_d^i : A^* \to A_d^*$ be the function defined for any word $u = a_0 a_1 \cdots a_n \in A^*$ by $\alpha_d^i(u) = (a_0, i)(a_1, i + 1 \bmod d)\cdots(a_n, i + n \bmod d)$. We simply denote α_d^0 by α_d and the word $\alpha_d(u)$ is called the well-formed word attached *to u.*

Note that the restriction of π_d to the set of well-formed words is one-to-one. For instance, the enriched word $(a,0)(b,1)(b,2)(a,0)$ is a well-formed word for $d = 3$. It is the unique well-formed word having the word *abba* as underlying word. Finally, given a language L, we write $L_d = \pi_d^{-1}(L) \cap \mathrm{K}_d$.

3.2 A first transfer theorem

Using the enriched alphabet and the well-formed words, the next theorem links a fragment with its enrichment by congruences modulo one integer. It transfers the expressiveness of modular predicates to the alphabet. An aware reader could notice that it is very similar to the wreath product principle of varieties. It is in fact not a coincidence, since this operation matches with a wreath product by the *length-multiplying* variety **MOD** (see Chaubard et al. (2006) for more details).

Theorem 3. *Let $\mathcal{F}[\sigma]$ be a fragment of logic, L a regular language and d a positive integer. Then the following properties are equivalent:*

(1) *L is definable by a formula of $\mathcal{F}[\sigma, \mathrm{MOD}^d]$,*
(2) *there exists some languages $L_0, \ldots, L_{d-1}$ of $\mathcal{F}[\sigma]$ over A_d^* such that:*

$$L = \bigcup_{i=0}^{d-1} \big((A^d)^* A^i \cap \pi_d(L_i \cap K_d)\big) \tag{a}$$

To prove the result, we need an auxiliary result which gives a decomposition of the language defined by a formula into smaller pieces.

Lemma 4. *Let $\mathcal{F}[\sigma, \mathrm{MOD}]$ be a fragment of logic and φ a formula of $\mathcal{F}[\sigma, \mathrm{MOD}^d]$. Then there exists d formulas ψ_i of $\mathcal{F}[\sigma, \mathrm{MOD}^d]$ that do not contain any predicate D_j^d and such that*

$$\varphi \equiv \bigvee_{i=0}^{d-1} (\psi_i \wedge D_i^d).$$

Moreover, we have:

$$L(\varphi) = \bigcup_{i=0}^{d-1} \big((A^d)^* A^i \cap L(\psi_i)\big).$$

Proof: For $i < d$, we define the formula ψ_i to be the formula φ where we replaced every predicate D_i^d by *true* and every D_j^d with $j \neq i$ by *false*. One should notice that, by definition of a fragment, the formulas ψ_i are in $\mathcal{F}[\sigma, \mathrm{MOD}^d]$. We can conclude the proof since the formula (D_i^d) recognizes the language $(A^d)^* A^i$. □

Proof of theorem 3: Let φ be a formula of $\mathcal{F}[\sigma, \mathrm{MOD}]$. Then φ belongs to $\mathcal{F}[\sigma, \mathrm{MOD}^d]$ for some $d > 0$. Using Lemma 4, we know it is sufficient to consider a formula φ without any length predicate. We transform it into a formula ψ by doing the following transformation:

$$\mathrm{MOD}_i^d(x) \text{ is replaced by } \bigvee_{a \in A} (\mathbf{a}, \mathbf{i})(x),$$
$$\mathbf{a}(x) \text{ is replaced by } \bigvee_{0 \leqslant i < d} (\mathbf{a}, \mathbf{i})(x).$$

The resulting formula ψ is in $\mathcal{F}[\sigma](A_d^*)$ and $L(\varphi) = \pi_d(L(\psi) \cap K_d)$. Conversely, we transform a formula ψ of $\mathcal{F}[\sigma](A_d^*)$ into a formula φ of $\mathcal{F}[\sigma, \mathrm{MOD}^d]$ by replacing every predicate $(\mathbf{a}, \mathbf{i})(x)$ in ψ by $\mathbf{a}(x) \wedge \mathrm{MOD}_i^d(x)$. We also get $L(\varphi) = \pi_d(L(\psi) \cap K_d)$. □

The previous theorem provides a semantic counterpart to the action of adding modular predicates to a fragment of logic. In the case where the fragment is *expressive enough*, this counterpart provides a transfer of decidability, as stated in the next corollary.

Corollary 5 (The transfer result)**.** *Let $\mathcal{F}[\sigma]$ be a fragment of logic. If $\mathcal{F}[\sigma]$ is decidable and if both K_d and **max** are definable in $\mathcal{F}[\sigma]$, then $\mathcal{F}[\sigma, \mathrm{MOD}^d]$ is decidable.*

Proof: The result comes from the fact that if **max** is definable, then using modular predicates the languages $(A^d)^* A^i$ are definable in $\mathcal{F}[\sigma, \mathrm{MOD}^d]$. If furthermore we can define the language of well-formed words, then item 2 of Theorem 3 is equivalent to the language L_d being definable in

$\mathcal{F}[\sigma]$ over the enriched alphabet. This language being computable from L, we get decidability. □

Remark: Corollary 5 applies to fragments $\mathcal{B}\mathbf{\Sigma}_k[\sigma]$, $\mathbf{FO}[\sigma]$, when $k \geqslant 2$ and σ contains either +1 or the order. It also applies to fragments $\mathcal{B}\mathbf{\Sigma}_1[\sigma]$, $\mathbf{FO}^2_k[\sigma]$, or $\mathbf{FO}^2[\sigma]$ when +1 is contained in σ.

4 Main results

As stated in the previous section, any language defined by a fragment with modular predicates can be done so with a formula using only congruences to one specific integer. In fact, there exists an infinite number of such witnesses. The remaining of the article is dedicated to the problem of deciding one witness, given a language. We call it the *delay problem* and can be explicitly stated as follows:
The delay question: Given a regular language L and a fragment $\mathcal{F}[\sigma]$, is it possible to compute an integer d such that L belongs to $\mathcal{F}[\sigma, \mathrm{MOD}]$ *if, and only if*, it belongs to $\mathcal{F}[\sigma, \mathrm{MOD}^d]$?

Remark that such an integer d could depend of L and $\mathcal{F}[\sigma]$. The denomination stems from the Delay Theorem of Straubing (1985) that solves a similar question for the enrichment by the successor predicate. Section 5 is devoted to solve the delay problem for different classes of varieties. It relies heavily on algebraic notions, in particular the framework of categories. We present here the main applications to fragments of logic, which are summed up in Figure 2. This figure does not include decidability of the smaller fragments of **FO** equipped with modular predicates: $\mathbf{FO}[\mathrm{MOD}]$ (Theorem 6), $\mathbf{FO}[+1, \mathrm{MOD}]$ (Theorem 7), $\mathbf{FO}[=, \mathrm{MOD}]$ (Theorem 8) and $\mathbf{FO}[=, +1, \mathrm{MOD}]$ (Theorem 9).

The first decidability results comes from the local property. Although it does not bring many new results, mainly reproving Barrington et al. (1992) and Dartois and Paperman (2013), it gives a unified proof for these fragments. *Local* varieties have a particular role in the previous work of Straubing, where they are identified as varieties that *behave* gently compared toward +1. In the context of modular predicates, they also have this good property that allows us to state a fairly generic statement under this assumption. A formal definition of locality can be found in Section 5.2.

Theorem 6 (Local case, for monoids varieties)**.** *Let $\mathcal{F}[\sigma]$ be a fragment equivalent to a local variety $\mathbf{V}$. Now let L be a regular language and s its stability index, then the following statements are equivalent.*

- *L belongs to $\mathcal{F}[\sigma, \mathrm{MOD}]$.*
- *L belongs to $\mathcal{F}[\sigma, \mathrm{MOD}^s]$.*
- *the stable monoid of L belongs to $\mathbf{V}$.*

Furthermore, if $\mathcal{F}[\sigma]$ is decidable, then so is $\mathcal{F}[\sigma, \mathrm{MOD}]$.

Example of interest includes $\mathbf{FO}^1[\varnothing]$, $\mathbf{FO}^2[<]$ or $\mathbf{FO}[<]$, which are equivalent to $\mathbf{J}_1$, **DA** and **A** respectively. The locality of $\mathbf{J}_1$ and **A** can be found in the article of Tilson (1987), the locality of **DA** is slightly more intricate (see Almeida (1996)).

	[<]	[<, MOD]	[Reg]
$\mathcal{B}\mathbf{\Sigma}_1 = \mathbf{FO}^2_1$	Simon (1975) Thomas (1982)	Chaubard et al. (2006)	Maciel et al. (2000)
$\mathbf{FO}^2_k$	Krebs and Straubing (2012) Kufleitner and Weil (2012)	**New result**	**New result**
$\mathbf{FO}^2$	Thérien and Wilke (1999)	Dartois and Paperman (2013)	**New result**
$\mathcal{B}\mathbf{\Sigma}_2$	Place and Zeitoun (2014)	**New result**	**New result**
$\mathcal{B}\mathbf{\Sigma}_k$	Open	Reduces to [<] **New result**	Reduces to [<] **New result**
$\mathbf{FO}$	McNaughton and Papert (1971) Schützenberger (1965)	Straubing (1994)	Barrington et al. (1992)

Fig. 2: Decidability results of first-order fragments

When the initial variety is local, we can nest our approach with the one with the successor predicates. It is no longer needed to use the intricate framework of categories since in this case, we can apply Corollary 5 to slightly simplify the question.

Theorem 7 (Local case, for semigroups varieties)**.** *Let $\mathcal{F}[\sigma]$ be a fragment corresponding to a local variety $\mathbf{V}$ Now let L be a regular language and s its stability index, then the following statements are equivalent.*

- *L belongs to $\mathcal{F}[\sigma, +1, \mathrm{MOD}]$.*
- *L belongs to $\mathcal{F}[\sigma, +1, \mathrm{MOD}^s]$.*
- *the local monoids of the stable semigroups belongs to $\mathbf{V}$.*

Furthermore, if $\mathcal{F}[\sigma]$ is decidable, then so is $\mathcal{F}[\sigma, +1, \mathrm{MOD}]$.

This theorem is a consequence of Proposition 29. Note that both $\mathbf{FO}[+1, \mathrm{MOD}]$ and $\mathbf{FO}^2[<, +1, \mathrm{MOD}]$ fall into the scope of this theorem. In the case of full first order logic, the successor predicate being definable with the order, the expressiveness remains unchanged. The reduction to logic of these results can be found in Subsection 5.4 for Theorem 7 and Subsection 5.2 for Theorem 6. Note that this provides decidability.

A generalized approach of the previous results brings fresh ones, although we fail to obtain a delay independent from the fragment. We need to assume some properties on the *varieties of categories* generated by the initial variety. In particular, we assume that the *path-equations* of the so called *global* of a variety use a bounded number of vertices. Under this assumption we successfully compute a delay.

Theorem 8 (Finite rank case)**.** *Let $\mathcal{F}[\sigma]$ be a fragment corresponding to a variety $\mathbf{V}$ of rank k. Now let L be a regular language and s its stability index, then the following statements are equivalent.*

- *L belongs to $\mathcal{F}[\sigma, \mathrm{MOD}]$.*
- *L belongs to $\mathcal{F}[\sigma, \mathrm{MOD}^{ks}]$.*

Furthermore, if $\mathcal{F}[\sigma, +1]$ is decidable, then so is $\mathcal{F}[\sigma, \mathrm{MOD}]$.

Example of application of this theorem include $\mathbf{FO}[=]$ which is known to be equivalent to the variety of rank 2 of aperiodic and commutative monoids, as well as the alternation hierarchy of $\mathbf{FO}^2[<]$ whose k^{th} level is of rank $2k$. This approach is detailed in Section 5.3. In those cases, this last theorem also provides decidability by reducing to decidability of the fragment with the successor predicate.

Finally, the next theorem provides a delay for all fragments containing the successor predicates. In particular, it reduces the decidability of $\mathcal{F}[\mathrm{Reg}]$ to the decidability of $\mathcal{F}[<, +1]$ providing decidability for the fragment $\mathbf{FO}^2_k[\mathrm{Reg}]$ and a reduction of the decidability of $\mathcal{B}\mathbf{\Sigma}_k[\mathrm{Reg}]$ to the decidability of $\mathcal{B}\mathbf{\Sigma}_k[<, +1]$, which itself reduces to decidability of $\mathcal{B}\mathbf{\Sigma}_k[<]$ thanks to Straubing (1985).

Theorem 9 (Infinitely testable case)**.** *Let $\mathcal{F}[\sigma]$ be a fragment corresponding to a variety $\mathbf{V}$ which is not a variety of groups. Now let L be a regular language and s its stability index, then the following statements are equivalent.*

- *L belongs to $\mathcal{F}[\sigma, +1, \mathrm{MOD}]$.*
- *L belongs to $\mathcal{F}[\sigma, +1, \mathrm{MOD}^s]$.*

Furthermore, if $\mathcal{F}[\sigma, +1]$ is decidable, then so does $\mathcal{F}[\sigma, +1, \mathrm{MOD}]$.

The condition that $\mathcal{F}[\sigma]$ is not equivalent to a group variety is necessary to apply the simplification of Corollary 5. However, in the case where $\mathcal{F}[\sigma]$ is indeed a variety of groups, then both $\mathcal{F}[\sigma, +1]$ and $\mathcal{F}[\sigma, \mathrm{MOD}]$ are decidable since varieties of groups are known to be local as variety of monoids but seems intricate when both +1 and MOD are in the signature since groups are not local as varieties of semigroups (for instance see book (Rhodes and Steinberg, 2009, page 104)). The proof of this last theorem is given in Subsection 5.4.

5 Solving the Delay problem

This section is devoted to solve the delay question for different classes of varieties.

We first present the framework of finite categories as well as some known results, and use it to reduce the combinatoric characterisation of Theorem 3 to the decidability of the global of a variety, an algebraic notion from the framework of finite categories.

The remainder of the section then uses this characterisation to solve the delay question for different classes of varieties. The first case is the simplest one of local varieties, where we get a clear characterisation of $\mathcal{F}[\sigma, \mathrm{MOD}]$. The second case, the finite rank, is a generalisation of the local case, where an algebraic characterisation of the global is known. Finally, the last case

solves the delay for a class of varieties where little is known about the global. It is the class of varieties of semigroups *expressive enough* and satisfying an extra property: the *infinitely testable property*, which is a new notion.

5.1 A derived category theorem

Finite categories: a short introduction. In this section, we present the theory of finite categories, as an extension of finite monoids. Informally, a category can be seen as a partial monoid where only some products are allowed. Nonetheless, notions from monoids can be correctly lifted, and we will consider varieties of categories. The framework of variety of categories has been successful to obtain algebraic characterizations of *wreath products* of varieties (Tilson, 1987). For example, the enrichment by modular predicates can be seen as a wreath product by a variety of morphisms. This comes from an adapted version of the Wreath Product Principle, that is evoked by Chaubard et al. (2006). We chose not to focus on this, since it would require to introduce additional definitions and proofs that are not necessary and would burden the article.

A graph X is a set of *objects* denoted $\mathrm{Ob}(X)$ such that for any couple of objects $(x, y) \in \mathrm{Ob}(X)$, we associate a set $X(x, y)$ of *arrows* from x to y. Two arrows e, f are coterminal if there exists $x, y \in \mathrm{Ob}(X)$ such that $e, f \in X(x, y)$. They are consecutive if there exists $x, y, z \in \mathrm{Ob}(X)$ such that $e \in X(x, y)$ and $f \in X(y, z)$. An arrow e is a loop from x if $e \in X(x, x)$. A composition law associates to each pairs of consecutive arrows, e, f an arrow ef. This law is said to be associative if for any consecutive arrows e, f, g we have $(ef)g = e(fg)$.

A *category* C is a graph with an associative composition law and containing for each object x an identity denoted 1_x. Thus the set of loops around a given object, equipped with the composition law, forms a monoid, called the *local monoid* of that object. Note that the terminology of local monoids of a category clashes with the terminology of local monoids of a semigroup. In fact, the two coincide when we consider the idempotent category of a semigroup, which is defined later.

Here we only consider categories as a generalization of finite monoids, since a monoid can be viewed as a one-object category. A *morphism of categories* $\eta : C \to D$ is an application $\eta : \mathrm{Ob}(C) \to \mathrm{Ob}(D)$ and for each pairs of object $(x, y) \in \mathrm{Ob}(C)$, an application $\eta : C(x, y) \to D(\eta(x), \eta(y))$ such that

(1) for any consecutive arrows e, f we have $\eta(ef) = \eta(e)\eta(f)$,

(2) for any $x \in \mathrm{Ob}(C)$, $\eta(1_x) = 1_{\eta(x)}$.

A *division* of categories $\tau : C \to D$ is given by a mapping $\tau : \mathrm{Ob}(C) \to \mathrm{Ob}(D)$, and for each pair of objects e and f, by a relation $\tau : C(e, f) \to D(\tau(e), \tau(f))$ such that

(1) $\tau(x)\tau(y) \subseteq \tau(xy)$ for consecutive arrows x, y,

(2) $\tau(x) \neq \varnothing$ for any arrow x,

(3) $1_{\tau(e)} \in \tau(1_e)$ for any object e of C.

We remark that the inverse of an onto morphism of categories is a division of a categories (but the converse is not true). Then a *variety of categories* is a class of categories closed under direct product and division.

Definition 10. *Given a variety of monoids* $\mathbf{V}$*, the* global *of* $\mathbf{V}$*, denoted* $\mathbf{gV}$*, is the class of all categories that divide a monoid of* $\mathbf{V}$*, when seen as a one-object category.*

Remark: Since the division of categories is a partial order and a variety is closed under product, the class of categories $\mathbf{gV}$ is closed by division and by product, and it is therefore a variety of categories.

Definition 11 (Consolidated semigroup, consolidated stamp)**.** *Let C be a finite category and $Arr(C)$ the set of arrows of C. We denote by C_{cd} the semigroup defined on the set*

$$E = Arr(C) \cup \{0\}$$

with for any $x \in E$, $0x = x0 = 0$, and for $x, y \in Arr(C)$,

$$x.y = \begin{cases} xy \text{ if } x \text{ and } y \text{ are consecutives arrows,} \\ 0 \text{ otherwise.} \end{cases}$$

The following proposition is a well-known result stating that the membership of a category in $\mathbf{gV}$ reduces to the membership of $\mathbf{V}$ is the variety is *expressive enough*. This is a *category* version of Corollary 5 which means that the membership of a language to an expressive enough fragment enriched with a finite set of modular predicates reduces to the membership of a different language to the fragment without them.

Background: the local predicates and derived category for locally testable language. In this section, we recall some known results that we will be using in the remainder of the article and give some intuitions about their significance. We first give the definition of the derived category for definite languages and provide the delay theorem of Straubing (1985) as well as its improvement by Tilson (1987). Let S be a semigroup, n an integer and $\eta : S^+ \to S$ the canonical semigroup morphism of S. The *n-derived category* of S with respect to definite languages, denoted $D_n(S)$, is the category with $S^{\leqslant n}$ as set of objects, and the arrows from u to v are the elements s of S such that there exists a word $w \in S^+$ that $\eta(w) = s$ and the suffix of size n of uw is equal to v. The n-derived category with respect to definite languages, of a regular language L, denoted $D_n(L)$, is the category $D_n(\eta_L(A^+))$. Finally we also introduce the *idempotents' category* of a semigroup S, denoted by S_E and defined by Tilson (1987) as follows. Its set of objects are the idempotents of S. And for e and f two idempotents, we set $S_E(e, f) = eSf$. We do not recall the definition of the wreath product of a variety $\mathbf{V}$ by $\mathbf{D}$, denoted by $\mathbf{V} * \mathbf{D}$. However, as our only use of this product is given by the following theorem, an unfamiliar reader can take the following theorem as a definition.

Theorem 12 (Delay theorem for definite languages)**.** *Let $\mathbf{V}$ be a variety and S a semigroup. The following conditions are equivalent.*

(1) *The semigroup S belongs to $\mathbf{V} * \mathbf{D}$.*
(2) *There exists an integer n such that $D_n(S)$ belongs to $\mathbf{gV}$.*
(3) *For $n = |S|$, $D_n(S)$ belongs to $\mathbf{gV}$.*
(4) *The category S_E belongs to $\mathbf{gV}$.*

For sufficiently expressive fragments, the operation of adding the local predicates corresponds to mapping the equivalent variety $\mathbf{V}$ to $\mathbf{V} * \mathbf{D}$. In fact, it will not be the case only if the fragment cannot use these predicates properly. In all cases, it is equivalent to adding the descriptive

local predicates defined in Section 2. The proof of the following proposition follows the proof of Theorem 3, by using an adapted notion of enriched alphabet. We omit the proof, that could be find in Paperman (2014).

Proposition 13. *Let $\mathcal{F}[\sigma]$ be a fragment of logic equivalent to a variety $\mathbf{V}$ and L a regular language with S as syntactic semigroup. The following conditions are equivalent.*

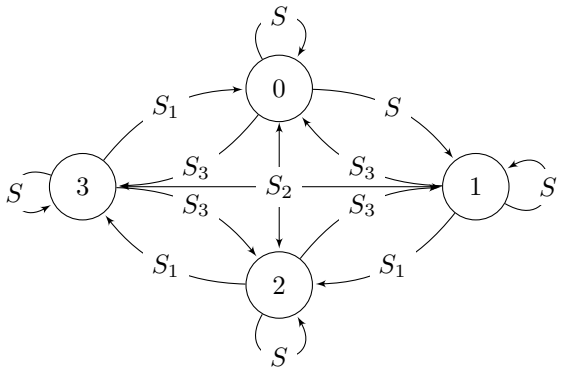
(1) *L is definable in $\mathcal{F}[\sigma, +1]$.*
(2) *S belongs to $\mathbf{V} * \mathbf{D}$.*
(3) *S_E belongs to $\mathbf{gV}$.*

The derived category relatively to modular languages. Following the preceding paragraph, we give the definition of the *derived category* adapted to modular languages which was largely inspired by the article of Chaubard et al. (2006).

Let $\varphi : A^* \to M$ be a morphism and d an integer. The *d-derived category* of φ, denoted $C_d(\varphi)$, is the category with $\mathbb{Z}_d$ as set of objects, and the arrows from i to j are the elements m of M such that there exists a word u satisfying $\varphi(u) = m$ and $i + |u| \equiv j \bmod d$. The d-derived category of a regular language L, denoted $C_d(L)$, is the category $C_d(\eta_L)$. The following lemma is a straightforward consequence of the definition that will be of some use.

Lemma 14. *Let d be a positive integer, and L be a regular language of stability index s. Then the local monoids of $C_d(L)$ are isomorphic to $\eta_L((A^d)^*)$. In particular, the local monoids of $C_s(L)$ are isomorphic to the stable monoid of L.*

Example 5. *The 4-derived category of the language $(aa)^*ab(bb)^*$ is given below. Let η be its syntactic morphism and S its stable monoid. Its stability index is 4.*



for $0 \leqslant i \leqslant 3$, $S_i = \eta((A^s)^* A^i)$

$S_0 = S = \{1, aa, bb, aabb, ba\}$

$S_1 = S_3 = \{a, b, aab, abb\}$

$S_2 = \{aa, bb, aabb, ba\}$

Proposition 15. *Let L be a regular language. For any $0 < d \leqslant d'$, if d divide d', then $C_{d'}(L)$ divides $C_d(L)$.*

Proof: Let L be regular language and $0 \leqslant d < d'$ be integers such that d divides d'. We define the relation $\tau : C_{d'}(L) \to C_d(L)$. The object application $\mathrm{Ob}(\tau) : \mathbb{Z}_{d'} \to \mathbb{Z}_d$ is defined by $\mathrm{Ob}(\tau)(x) = x \bmod d$ for any $x \in \mathbb{Z}_{d'}$. Let (x, m, y) be an arrow of $C_{d'}(L)$. By definition, there exists $u \in A^*$ such that $\eta_L(u) = m$ and $|u| \equiv y - x \bmod d'$. Let $a = x \bmod d$ and $b = y \bmod d$.

Then, since d divides d', $|u| \equiv b - a \bmod d$. Thus, the arrow (a, m, b) is in $C_d(L)$. We define $\tau(x, m, y) = (a, m, b)$. The application τ is a morphism and for any $(x, m, y) \neq (x, m', y)$, we have $\tau(x, m, y) \neq \tau(x, m', y)$. Therefore, τ defines a division from $C_{d'}$ to C_d. □

The derived category theorem was originally proved by Tilson (1987) for varieties of monoids and semigroups. Unfortunately the case of modular languages can not be dealt with the framework of Tilson since they do not form a variety of language. However it has been extended to *length-multiplying* varieties in the PhD thesis of Chaubard (2007). Since this work is only available in french, we provide a proof inspired by the work of Chaubard, but adapted to our framework.

Theorem 16. *Let $\mathcal{F}[\sigma]$ be a fragment of logic equivalent to a variety of monoids $\mathbf{V}$, L a regular language and d a positive integer. Then the following properties are equivalent:*

(1) *L is definable by a formula of $\mathcal{F}[\sigma, \mathrm{MOD}^d]$,*

(2) *there exists some languages $L_0, \ldots, L_{d-1}$ of $\mathcal{F}[\sigma]$ over A_d^* such that:*

$$L = \bigcup_{i=0}^{d-1} \left((A^d)^* A^i \cap \pi_d(L_i \cap K_d) \right) \tag{a}$$

(3) *the category $C_d(L)$ belongs to $\mathbf{gV}$.*

Proof: The equivalence between the two first points is obtained directly by Theorem 3. We only prove the equivalence between (3) and (2). As always, we denote by $\eta_L : A^* \to M_L$ the syntactic morphism of L and $P = \eta_L(L)$ its accepting set.

(3) → (2): Assume that $C_d(L)$ belongs to $\mathbf{gV}$. By definition, it means that there exists a division of categories $\tau : C_d(L) \to M$, where M is a monoid of $\mathbf{V}$ seen as a one object category. We need to define some appropriate languages L_i for $0 \leqslant i < d$. To this end, we construct an *adequate* morphism from A_d^* to M.
Let then $\beta : A_d^* \to M$ be defined by $\beta(a, i) = m$ where m is any element in $\tau(i, \eta(a), i + 1 \bmod d)$. For $0 \leqslant i < d$, let $E_i = \bigcup_{m \in P} \tau(0, m, i)$ and $L_i = \beta^{-1}(E_i)$. Because M is in $\mathbf{V}$, these languages are all in $\mathcal{F}[\sigma]$.
It remains to verify that these languages satisfy the hypothesis. This is equivalent to check that for all $i < d$

$$\alpha_d(L \cap (A^d)^* A^i) \subseteq L_i \text{ and } \alpha_d(L^c \cap (A^d)^* A^i) \cap L_i = \varnothing.$$

Let $u = (a_0, 0) \cdots (a_n, p)$ be a well-formed word of A_d^*, by construction of β, we have

$$\beta(u) = m = m_1 \cdots m_n \in \tau(0, \eta_L(a_1), 1) \cdots \tau(p, \eta_L(a_n), p + 1) \subseteq \tau(0, \eta_L(a_1 \cdots a_n), p + 1)$$

Therefore, we have

$$\beta\Big(\alpha_d(L \cap (A^d)^* A^i)\Big) \subseteq E_i.$$

Furthermore, since τ is a division, for all $u \in \alpha_d(L^c \cap (A^d)^* A^i)$, $\beta(u) \notin \tau(0, m, i)$ for all $m \in P$ and thus $\beta(u) \notin E_i$.

(2) → (3): Let $L_0, \dots, L_{d-1}$ be languages of $\mathcal{F}[\sigma]$ as stated by (2). Then each of them is definable by a monoid of $\mathbf{V}$, and since varieties are closed by product, there exists a morphism $\beta : A_d^* \to M$ that recognizes them all, with $M \in \mathbf{V}$. We now prove that $C_d(L)$ divides M. Let $\tau : C_d(L) \to M$ be defined by

$$\tau(i,x,j) = \{\beta(u) \mid \exists u \in K_d(i,j) \text{ s.t. } \eta_L(\pi_d(u)) = x\}$$

The application τ satisfies the first three axioms of a division of categories.

(1) We have $1 \in \tau(i,1,i)$ for any i of $\mathbb{Z}/d\mathbb{Z}$.

(2) Let (i,x,j) be an arrow of $C_d(L)$. By definition, there exists v in $(A^d)^* A^{j-i}$ such that $\eta_L(v) = x$. Let $u = \alpha_d^i(v) \in K_d(i,j)$. By definition, $\beta(u) \in \tau(i,x,j)$ and thus $\tau(i,x,j) \neq \varnothing$.

(3) Let (i,x,j) and (j,x',k) be two arrows in $C_d(L)$ and $m \in \tau(i,x,j)$, $m' \in \tau(j,x',k)$. By hypothesis, there exists $u \in K_d(i,j)$ and $u' \in K_d(j,k)$ such that $\beta(u) = m$ and $\beta(u') = m'$, and such that $\eta_L(\pi_d(u)) = x$ and $\eta_L(\pi_d(u')) = x'$. Then, mm' belongs to $\tau(i,xx',k)$ since $mm' = \beta(uu')$, $uu' \in K_d(i,k)$ and $\eta_L(\pi_d(uu')) = xx'$.

Unfortunately, it could happen that τ does not satisfy the last condition. Without detailing, this is due to the fact that some elements of the syntactic congruence of L might merge when appearing at some specific congruences, leading to non empty intersection of images of arrows. In the following, we use the idea that for any pair of elements there exists a congruence that separates them by definition of the syntactic congruence.
Thus, we now introduce a *twisted product* of τ, denoted by $\otimes_d\tau : C_d(L) \to M^d$ and formally define it by

$$\otimes_d\tau(i,x,j) = \big(\tau(i,x,j), \tau(i+1,x,j+1), \dots, \tau(i+d-1,x,j+d-1)\big)$$

Because $\otimes_d\tau$ is a product of τ by it self d times, it satisfies immediately the first three axioms of a division of categories. We now prove that $\otimes_d\tau$ is a division by proving the separation axiom.

(4) Let x, x' be two distinct elements of M_L such that (i,x,j) and (i,x',j) are arrows of $C_d(L)$. We first prove that there exists r,t satisfying $r - t = j - i$ and such that $\tau(t,x,r) \cap \tau(t,x',r) = \varnothing$ and then conclude by using $\otimes_d\tau$. Let v and v' in $(A^d)^* A^{j-i}$ such that $\eta_L(v) = x$ and $\eta_L(v') = x'$.
Since $x \neq x'$, and by definition of M_L, the syntactic monoid of L, we can assume that there exists $p, q \in A^*$ such that $pvq \in L$ if and only if $pv'q \notin L$. Let $y = \eta_L(pvq)$ and $y' = \eta_L(pv'q)$. We remark that $(0,y,k)$ and $(0,y',k)$ are arrows $C_d(L)$ for $k = |pvq| \bmod d = |pv'q| \bmod d$. Without loss of generality, we assume pvq to be in L, the other case being symmetrical. By hypothesis, we have the following:

$$\eta_L^{-1}(y) \cap (A^d)^* A^k \subseteq L_k$$
$$\eta_L^{-1}(y') \cap (A^d)^* A^k \cap L_k = \varnothing$$

However

$$\tau(0,y,k) = \beta \circ \alpha_d\big(\eta_L^{-1}(y) \cap (A^d)^* A^k\big)$$
$$\tau(0,y',k) = \beta \circ \alpha_d\big(\eta_L^{-1}(y') \cap (A^d)^* A^k\big)$$

Since L_k is recognized by M through the morphism β we have

$$\tau(0, y, k) \cap \tau(0, y, k) = \varnothing.$$

To conclude, it suffices to notice that

$$\tau(0, s, t) \cdot \Big(\tau(t, x, r) \cap \tau(t, x', r)\Big) \cdot \tau(r, t, k) \subseteq \tau(0, y, k) \cap \tau(0, y, k) = \varnothing,$$

where $s = \eta_L(p)$, $t = |p|$, $r = t + j - i$ and $t = \eta_L(q)$. Since both $\tau(0, s, i)$ and $\tau(j, t, k)$ are nonempty, we conclude that $\tau(t, x, r) \cap \tau(t, x', r) = \varnothing$. We proved that for every arrow (i, x, j) and (i, x', j) in $C_d(L)$, there exists r, t $r - t = j - i$ and such that $\tau(t, x, r) \cap \tau(t, x, r) = \varnothing$. Therefore, we obtain that $\otimes_d \tau(i, x, j) \cap \otimes_d \tau(i, x', j) = \varnothing$ for every coterminal arrows (i, x, j) and (i, x', j) in $C_d(L)$, which concludes the proof. □

5.2 local case

For any variety $\mathbf{V}$, we define $\mathbf{QV}$ to be the class of morphisms (lm-variety of morphisms to be precise, see the article of Pin and Straubing (2005) for more details) whose stable monoid is in $\mathbf{V}$. Following the article of Tilson (1987), we denote by $\ell\mathbf{V}$ the variety of categories whose local monoids are all in $\mathbf{V}$. A variety of monoids $\mathbf{V}$ is said to be *local* if $\mathbf{gV} = \ell\mathbf{V}$. The next theorem makes explicit the link between $\mathbf{QV}$ and $\ell\mathbf{V}$.

Theorem 17. *Let $\mathbf{V}$ be a variety and L a regular language of A^* of stability index s. The following properties are equivalent:*

(1) *L is recognized by a morphism in $\mathbf{QV}$,*

(2) *there exists an integer d such that $C_d(L)$ is in $\ell\mathbf{V}$,*

(3) *$C_s(L)$ is in $\ell\mathbf{V}$.*

Proof:
$1 \to 3$. If L is recognized by a stamp in $\mathbf{QV}$, then its syntactic stamp is also in $\mathbf{QV}$ and its stable monoid is in $\mathbf{V}$. But, thanks to Lemma 14, the local monoids of $C_s(L)$ belong to $\mathbf{V}$, and thus $C_s(L)$ is in $\ell\mathbf{V}$.
$3 \to 2$. Is obvious.
$2 \to 1$. Suppose that $C_d(L)$ is in $\ell\mathbf{V}$. Then the local monoids of $C_d(L)$, which are isomorphic to $\eta_L((A^d)^*)$ by Lemma 14, belong to $\mathbf{V}$. Thus $\eta_L((A^{ds})^*)$, which is a submonoid of $\eta_L((A^d)^*)$, also belongs to $\mathbf{V}$. Finally, by definition of the stability index, the monoid $\eta_L((A^s)^*) = \eta_L((A^{ds})^*)$ is in $\mathbf{V}$ and thus η_L is in $\mathbf{QV}$. □

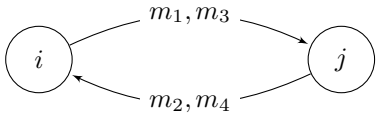
Observe that any monoid of $\mathbf{V}$, viewed as a one-object category, belongs to $\ell\mathbf{V}$. Therefore by definition of $\mathbf{gV}$, any category of $\mathbf{gV}$ divides a category of $\ell\mathbf{V}$, and thus $\mathbf{gV} \subseteq \ell\mathbf{V}$. The varieties satisfying $\mathbf{gV} = \ell\mathbf{V}$ are exactly the local varieties. Combining this with Theorem 16 and since the stability index and the stable monoid of a given regular language are computable, one gets the following corollary.

Corollary 18. *Let $\mathcal{F}[\sigma]$ be a fragment equivalent to a local variety $\mathbf{V}$. Then $\mathcal{F}[\sigma]$ is decidable if and only if $\mathcal{F}[\sigma, \mathrm{MOD}]$ is decidable. Furthermore, the fragment $\mathcal{F}[\sigma, \mathrm{MOD}]$ is equivalent to $\mathbf{QV}$.*

Adding modular predicates does not always coincide with the **Q** operation. A counterexample is the variety **J**, which is known to be nonlocal. Chaubard et al. (2006) proved the decidability of $\mathcal{B}\mathbf{\Sigma}_1[<, \mathrm{MOD}]$, using the characterization of $g\mathbf{J}$ given by Knast (1983) (see Figure 3). Using this characterization, we can prove that the language $(aa)^*ab(bb)^*$, whose stable monoid is in **J** does not satisfy Knast's equation since

$$\begin{aligned}(m_1m_2)^\omega(m_3m_4)^\omega = (aa)^\omega(bb)^\omega &= aabb \\ &\neq ab = (aa)^\omega ab(bb)^\omega = (m_1m_2)^\omega m_1m_4(m_3m_4)^\omega\end{aligned}$$

It is therefore not definable in $\mathcal{B}\mathbf{\Sigma}_1[<, \mathrm{MOD}]$ (see Example 5).



$$(m_1m_2)^\omega(m_3m_4)^\omega = (m_1m_2)^\omega m_1m_4(m_3m_4)^\omega$$

Fig. 3: Path equation of $g\mathbf{J}$ by Knast.

5.3 Finite rank

Although the local property gives a nice algebraic characterisation, it only applies to a few varieties. Nonetheless, we can still obtain a delay when the global is well-understood. To be more precise, we now prove a delay for varieties where equations for the global are known. As the global is a variety of categories, we first extend the framework of profinite equations to categories. Note finally that this is the only case where we obtain a delay that is greater than the stability index. The main applications on fragments of logic are given in Corollary 26.

Path equations The theory of profinite equation of varieties of monoids extends naturally to path equations on graphs, characterising varieties of categories. The complexity of a variety of categories is given by its rank, which is the minimal size required to describe the variety in terms of path equations. Let X be a graph and E the set of arrows of X. Then X^* is the set of words on $u = u_0 \cdots u_n \in E^*$ such that for all $i < n$, u_i and u_{i+1} are consecutive arrows. X^* is named the free category on X. Let u and v be coterminal paths of X^*. Then

$$r(u, v) = \min\left\{n \mid \exists \varphi : X^* \to C \text{ with } n = |C| \text{ and } \varphi(u) \neq \varphi(v)\right\}$$

where φ is a category morphism and C a finite category. We define $d(u, v) = 2^{-r(u,v)}$ which is an ultrametric distance on X^*. The completion of X for this metric is called the profinite free category on X and is denoted by $\widehat{X}^*$. The following proposition is very standard in the framework of (pseudo-)varieties of monoids and categories.

Proposition 19. *Let X be a graph, C a finite category and $\varphi : X^* \to C$ a morphism of categories. Then, there exists a unique continuous function $\widehat{\varphi} : \widehat{X^*} \to C$ that extends φ. Furthermore, for any $u \in \widehat{X^*}$, there exists $v \in X^*$ such that $\widehat{\varphi}(u) = \widehat{\varphi}(v)$.*

Let X be a graph and $u, v \in \widehat{X^*}$ coterminal profinite paths. We say that the finite category C satisfy the equation $(X, u = v)$ if for any morphism $\varphi : X^* \to C$, we have $\widehat{\varphi}(u) = \widehat{\varphi}(v)$.

Theorem 20 (Tilson)**.** *Every non trivial variety of finite categories is defined by a set of equations.*

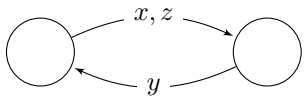
Definition 21 (Rank of a variety)**.** *We say that a variety of monoids* $\mathbf{V}$ *has a rank* k *if its global is defined by a set of bounded path equations with at most* k *vertices. If* $\mathbf{V}$ *has a finite rank, we denote by rank*$(\mathbf{V})$ *the minimal* k *such that* $\mathbf{V}$ *has a rank* k*.*

We remark that the varieties of rank one are exactly the local ones. Furthermore, most of the known fragments of logic are equivalent to a variety of finite rank. The question remains however open in some cases, as for instance for the levels of the dot-depth hierarchy.

Example 6. *We now give several varieties where equations for the global are known.*

1. *Several varieties are known to be local. For instance, the variety of semilattice monoids* $\mathbf{J}^1 = [\![xy = yx, x^2 = x]\!]$*, the variety* $\mathbf{DA} = [\![(xy)^\omega x (xy)^\omega = (xy)^\omega]\!]$*, the variety of aperiodic monoids* $\mathbf{A} = [\![x^\omega = x^{\omega+1}]\!]$*.*

2. *The variety of commutative monoids* $\mathbf{Com} = [\![xy = yx]\!]$*. The variety of categories* $g\mathbf{Com}$ *is defined below and thus is of rank* 2*.*

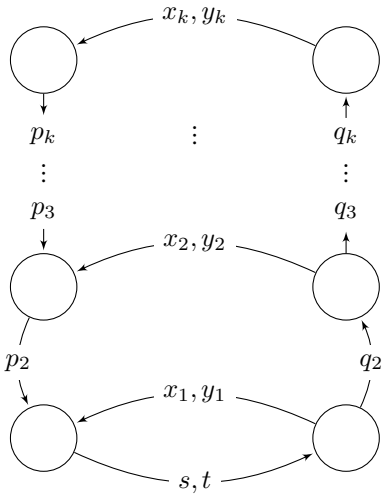


$xyz = zyx$

3. *A recent algebraic description of the languages definable by formulas of* $\mathcal{B}\mathbf{\Sigma}^2_{k+1}[<]$ *was established in Krebs and Straubing (2012); Kufleitner and Weil (2012). In the subsequent we will denote by* $\mathbf{V}_k$ *the equivalent variety of monoids. This result was extended to* $\mathcal{B}\mathbf{\Sigma}^2_{k+1}[<, \mathrm{LOC}]$ *in Kufleitner and Lauser (2012). From this latter result we derive the following description of* $\mathbf{gV}_k$*, giving a rank of at most* $2k$*.*

Theorem 22 (A Delay Theorem for finite rank varieties)**.** *Let* $\mathcal{F}[\sigma]$ *be a fragment equivalent to a variety* $\mathbf{V}$ *rank* k*. A language* L *belongs to* $\mathcal{F}[\sigma, \mathrm{MOD}]$ *if and only if* L *belongs to* $\mathcal{F}[\sigma, \mathrm{MOD}^{ks}]$*.*

Proof: First notice that since the if condition is trivial, we only need to prove the only if implication. Remark now that if $\mathrm{rank}(\mathbf{V}) = 1$ then the variety is local and we know that we can restrict to congruence modulo the stability index. For the rest of the proof we assume that $\mathrm{rank}(\mathbf{V}) = k > 1$.
Let now d be such that $C_d(L) \in \mathbf{gV}$. Without loss of generality, we assume that d is greater than k. Indeed if $d \leqslant k$, we consider $d' = dk$. Then by Proposition 15 $C_{d'}(L)$ divides $C_d(L)$ and thus also belongs to $\mathbf{gV}$.
So in the remainder of the proof we will assume that $ds > ks$. Since $C_d(L) \in \mathbf{gV}$ we know that $C_{ds}(L) \in \mathbf{gV}$. Then $C_{ds}(L)$ satisfies every path equation $(X, u = v)$ defining $\mathbf{gV}$. The goal of this proof is to show if $C_{ks}(L)$ does not satisfy a path equation defining $\mathbf{gV}$, then $C_{ds}(L)$ cannot satisfy it either.



We define $U_1 = (sx_1)^\omega s(y_1 t)^\omega$ and $V_1 = (sx_1)^\omega t(y_1 t)^\omega$

$$U_k = (p_k U_{k-1} q_k x_k)^\omega p_k U_{k-1} q_k (y_k p_k U_{k-1} q_k)^\omega$$

$$V_k = (p_k U_{k-1} q_k x_k)^\omega p_k V_{k-1} q_k (y_k p_k U_{k-1} q_k)^\omega$$

$g\mathbf{V}_k$ satisfies the equation $U_k = V_k$

Fig. 4: An equation for $\mathbf{gV}_k$.

So assume that there exists a path equation $(X, u = v)$ of rank k defining $\mathbf{gV}$ that is not satisfied by $C_{ks}(L)$. Then, there exists a category morphism $\varphi : X^* \to C_{ks}(L)$ such that $\widehat{\varphi}(u) \neq \widehat{\varphi}(v)$. We define $V = \varphi(\mathrm{Ob}(X))$ the set of objects of $C_{ks}(L)$ that have a preimage by φ, and

$$E = \{(i, m, j) \in C_{ks}(L) \mid \exists e \in X \ \varphi(e) = (i, m, j)\}$$

the set of arrows that have a preimage by φ. Notice that $E \subseteq V \times M_L \times V$.

We will construct a category morphism $\psi : X^* \to C_{ds}(L)$ such that $\widehat{\psi}(u) \neq \widehat{\psi}(v)$. In order to do that, we define a map $\theta : V \to C_{ds}(L)$ such that for all (i, m, j) in E, $(\theta(i), m, \theta(j))$ is an arrow of $C_{ds}(L)$.

Lemma 23. *There exists a smallest integer $i_V < ks$ such that $\{i_V+1, \ldots, i_V+s-1 \bmod ks\} \cap V = \varnothing$.*

Proof: As the size of X is k, the size of V is at most k. Then the maximal distance between two consecutive vertices of V is at least $ks/k = s$. □

We define $\theta : V \to \mathrm{Ob}(C_{ds}(L))$ as follow :

$$\theta : \begin{cases} i \mapsto i \bmod ds \text{ if } i \leqslant i_V \\ i \mapsto ds + i - ks \text{ otherwise.} \end{cases}$$

The idea behind this is that $i = \theta(i)$ if i appears before the gap and $ks - i = ds - \theta(i)$ if i appears after it. Then each arrow from E will either appear directly as it does for $C_{ks}(L)$ if it does not go over the gap, and since the gap is of size s, we will be able to pump the arrows that go over it.

Lemma 24. *For any arrow (i, m, j) of E, $(\theta(i), m, \theta(j))$ is an arrow of $C_{ds}(L)$.*

Proof: Let (i, m, j) be an arrow of E. Then there exists a word u such that $\eta_L(u) = m$ and $i + |u| = j \bmod ks$. We now distinguish the cases depending on the length of u.

- If $|u| \geqslant s$, then we know, by definition of the stability index, that for any positive integer ℓ, there exists a word u_ℓ such that $\ell s \leqslant |u_\ell| < (\ell+1)s$, $|u| = |u_\ell| \bmod s$ and $u \equiv_L u_\ell$. Then as θ preserves the congruence modulo s, $(\theta(i), \eta_L(u_\ell), \theta(j)) = (\theta(i), m, \theta(j))$ is an arrow of $C_{ds}(L)$.
- If $|u| < s$, then we have to treat several subcases:
  - If $\theta(i) = i$ and $\theta(j) = j$, then $\theta(i) + |u| = \theta(j) \bmod ds$. Thus $(\theta(i), m, \theta(j))$ is an arrow of $C_{ds}(L)$.
  - If $\theta(i) = ds + i - ks$ and $\theta(j) = ds + j - ks$, then as u has a size smaller than s, we have $i < j$ and $\theta(j) - \theta(i) = j - i$. Consequently $\theta(i) + |u| = \theta(j) \bmod ds$ and $(\theta(i), m, \theta(j))$ is an arrow of $C_{ds}(L)$.
  - If $\theta(i) = ds + i - ks$ and $\theta(j) = j$, then $i + |u| = j + ks$. So $\theta(i) + |u| = ds + i - ks + |u| = j + ds$. The same word u labels an arrow from $\theta(i)$ to j and thus $(\theta(i), m, \theta(j))$ is an arrow of $C_{ds}(L)$.
  - Finally, the case where $\theta(i) = i$ and $\theta(j) = ds + j - ks$ cannot happen since it implies that $i \leqslant i_V$ and $j > i_V + s$, and that $|u| = j - i > s \bmod ks$ which contradicts the $|u| < s$ hypothesis.

□

We now define a new morphism $\psi : X^* \to C_{ds}$. We proceed as follow:

- First we define $\mathrm{Ob}(\psi)$ to be $\theta \circ \mathrm{Ob}(\varphi)$.
- We now have to define ψ on arrows. Let e be an arrow of X and $\varphi(e) = (i, m, j)$. We set $\psi(x) = \big(\theta(i), m, \theta(j)\big)$. This is well defined thanks to Lemma 24.

Lemma 25. *Let u be a path in X^*. If $\varphi(u) = (i, m, j)$, then $\psi(u) = \big(\theta(i), m, \theta(j)\big)$.*

Proof: Let $u = u_1 \cdots u_n \in X^*$ such that $\varphi(u_\ell) = (i_\ell, m_\ell, j_\ell)$ and $\varphi(u) = (i, m, j)$. Therefore, $\psi(u_\ell) = \big(\theta(i_\ell), m_\ell, \theta(j_\ell)\big)$. However, since for all $1 \leqslant \ell < n$ $j_\ell = i_{\ell+1}$, we have $\phi(u) = (i_1, m_1 \cdots m_n, j_n) = (i, m, j)$ and $\psi(u) = \big(\theta(i_1), m_1 \cdots m_n, \theta(j_n)\big) = \big(\theta(i), m, \theta(j)\big)$. □

Recall that $\widehat{\varphi}(u) \neq \widehat{\varphi}(v)$. Then we can find $u', v' \in X^*$ co-terminal paths of X^* such that $\varphi(u') = \widehat{\varphi}(u)$, $\varphi(v') = \widehat{\varphi}(v)$, $\psi(u') = \widehat{\psi}(u)$ and $\psi(v') = \widehat{\psi}(v)$. We set $u' = u_1 \cdots u_n$ with $u_i \in X$ for any i and $v' = v_1 \cdots v_p$ with $v_i \in X$ for any i. To conclude we show that $\psi(u') \neq \psi(v')$ which is absurd. Indeed, if $\varphi(u') = (i, m, j) \in C_{ks}(L)$ and $\varphi(v') = (i, m', j) \in C_{ks}(L)$, then $m' \neq m$ since $\widehat{\varphi}$ separates u and v. Furthermore, by Lemma 25, we also have $\psi(u') = \big(\theta(i), m, \theta(j)\big)$ and $\psi(v') = \big(\theta(i), m', \theta(j)\big)$ in $C_{sd}(L)$. Finally $\psi(u') \neq \psi(v')$ and thus C_{ds} does not satisfy $(X, u = v)$, holding a contradiction.

□

Combining the previous theorem with the decidable path equations given in Example 6 yields the following corollaries.

Corollary 26. *Given a regular language L of stability index s and an integer $k > 0$. We have the following results.*

- *L belongs to $\mathbf{FO}^2_k[<, \mathrm{MOD}]$ if, and only if, it belongs to $\mathbf{FO}^2_k[<, \mathrm{MOD}^{2ks}]$.*
- *L belongs to $\mathbf{FO}[=, \mathrm{MOD}]$ if, and only if, it belongs to $\mathbf{FO}[=, \mathrm{MOD}^{2s}]$.*

As the corresponding global varieties are decidable, we get that the fragments $\mathbf{FO}[=, \mathrm{MOD}]$ and $\mathbf{FO}^2_k[<, \mathrm{MOD}]$ for any $k > 0$ are decidable.

5.4 Infinitely testable case

In this Section, we present the *infinitely testable* property. We then prove that for any expressive enough fragment equipped with all regular predicates, this property holds, leading to a delay. In fact, Proposition 27 proves that, given Proposition 13, as soon as a fragment contains the local predicates, it will be infinitely testable. Theorem 28 then proves that a delay can be computed in this latter case. Informally, a variety is infinitely testable if the membership of a language to the variety only depends on words *long enough*.

Definition. Given a semigroup S, the *idempotents' ideal* of S, denoted $\mathcal{I}_E(S)$, is the ideal of S generated by its idempotents, i.e. $\mathcal{I}_E(S) = SE(S)S$, where $E(S)$ denotes the set of idempotents of S. Note also that given a morphism $\eta : A^+ \to S$, it is the semigroup of all elements of S having an infinite number of preimages by η. An aware reader could notice that $\mathcal{I}_E(S)$ is the set of all elements of S that are $\mathcal{J}$-below an idempotent. A variety of semigroups $\mathbf{V}$ is said to be *infinitely testable* if the membership of a semigroup to $\mathbf{V}$ is equivalent to the membership of its idempotents' ideal. Informally, a variety is infinitely testable if its membership can be reduced to an algebraic condition on the idempotents' ideal. By extension, we say that a fragment of logic is infinitely testable if it is characterized by an infinitely testable variety.

Example 7. *The fragment $\mathbf{FO}[=]$ is equivalent to the aperiodic and commutative variety $\mathbf{ACom}$. This fragment is also described by the equations $xy = yx$ and $x^{\omega+1} = x^\omega$. This fragment is not infinitely testable. For instance the language equal to the singleton $\{ab\}$ has a trivial idempotents' ideal while it is not definable in $\mathbf{FO}[=]$.*

Example 8. *The fragment $\mathbf{FO}[+1]$ is equivalent to the languages whose syntactic semigroup belongs to the variety: $\mathbf{ACom} * \mathbf{D}$ (Straubing, 1994, Theorem VI.3.1). This fragment is also described by the profinite equation*

$$x^\omega u y^\omega v x^\omega w y^\omega = x^\omega w y^\omega v x^\omega u y^\omega \quad . \tag{a}$$

We now show that it is an infinitely testable fragment. Let L be a regular language and S its syntactic semigroup. We simply prove that if the equation (a) is not satisfied by S, then it is not satisfied by $\mathcal{I}_E(S)$. Suppose that there exists $x, y, u, v, w \in S$ such that the equation (a) is not satisfied. Then by setting: $x' = x^\omega$, $y' = y^\omega$, $u' = x^\omega u y^\omega$, $v' = y^\omega v x^\omega$, $w' = x^\omega w y^\omega$. All new variables belong to $\mathcal{I}_E(S)$ and they also fail to satisfy (a).

In fact, the approach given in the last example can be generalised to any variety of the form $\mathbf{V} * \mathbf{D}$. This is proved by Proposition 27.

Proposition 27. *Let $\mathbf{V}$ be a variety. The variety $\mathbf{V} * \mathbf{D}$ is infinitely testable.*

Proof: Let L be a regular language with $\eta : A^* \to M_L$ its syntactic morphism and $S = \eta_L(A^+)$ its syntactic semigroup. Using Theorem 12, we have that S belongs to $\mathbf{V} * \mathbf{D}$ if and only if S_E

belongs to $\mathbf{gV}$. To conclude, we just notice that by definition, $(\mathcal{I}_E(S))_E = S_E$, and therefore $\mathbf{V} * \mathbf{D}$ is infinitely testable. □

We finally prove here that if a fragment is equivalent to a variety whose global is infinitely testable, then we can effectively compute a delay, which furthermore is independent from the fragment. For varieties of the form $\mathbf{V} * \mathbf{D}$, this also gives the decidability thanks to Proposition 27 and Corollary 5.

Theorem 28. *Let $\mathcal{F}[\sigma]$ be a fragment equivalent to a variety $\mathbf{V}$ which is not a group variety and let L be a language of stability index s. Then L belongs to $\mathcal{F}[\sigma, +1, \mathrm{MOD}]$ if and only if L belongs to $\mathcal{F}[\sigma, +1, \mathrm{MOD}^s]$.*

Proof: First by Theorem 3, a language L belongs to $\mathcal{F}[\sigma, +1, \mathrm{MOD}]$ if and only if there exists $d > 0$, $L_0, \ldots, L_{d-1}$ in $\mathcal{F}[\sigma, +1]$ such that

$$L = \bigcup_{i<d} (A^d)^* A^i \cap \pi_d(L_i \cap K_d).$$

Because $\mathcal{F}[\sigma]$ is a fragment which is a variety of monoids but not a group variety, the language K_d and **max** belongs to $\mathcal{F}[\sigma, +1]$. We recall that $L_d = \pi_d^{-1}(L) \cap \mathrm{K}_d$ for any $d > 0$. Thus, for $i < d$, $L_i \cap K_d$ belongs to $\mathcal{F}[\sigma, +1]$ and we have the equality

$$L_= \bigcup L_i \cap A_d(0, i) \cap K_d \in \mathcal{F}[\sigma, +1].$$

Therefore, L belongs to $\mathcal{F}[\sigma, +1, \mathrm{MOD}]$ if and only if there exists d such that L_d belongs to $\mathcal{F}[\sigma, +1]$. Thus, it suffices to prove that if L_{ds} is definable in $\mathcal{F}[\sigma, +1]$, then L_s is in $\mathcal{F}[\sigma, +1]$ as well. We set $\eta_s : A_s^+ \to S_s$ and $\eta_{ds} : A_{ds}^+ \to S_{ds}$ the syntactic morphisms of L_s and L_{ds} respectively.
Claim. The semigroup $\mathcal{I}_E(S_s)$ divides $\mathcal{I}_E(S_{ds})$.

Before proving this claim, let us remark that since a variety of semigroups is closed by division, this claim ends the proof. Since if L belongs to $\mathcal{F}[\sigma, \mathrm{MOD}^{ds}]$ then S_{ds} belongs to $\mathbf{V} * \mathbf{D}$ and therefore $\mathcal{I}_E(S_{ds})$ belongs to $\mathbf{V} * \mathbf{D}$ as well. By division, $\mathcal{I}_E(S_s)$ belongs to $\mathbf{V} * \mathbf{D}$, and thanks to Proposition 27, S_s belongs to $\mathbf{V} * \mathbf{D}$. Finally, we deduce that L_s belongs to $\mathcal{F}[\sigma + 1]$.

We now aim to construct a division from $\mathcal{I}_E(S_s)$ to $\mathcal{I}_E(S_{ds})$. This is done through the enriched alphabet. We introduce the following projection

$$h : \begin{cases} A_{ds}^+ & \to & A_s^+ \\ (a, i) & \mapsto & (a, i \bmod s) \end{cases}$$

and F_d the language of *well-formed factors*, which is the set of well-formed words that do not necessarily start by a letter $(a, 0)$. Note that $L_{ds} = h^{-1}(L_s) \cap \mathrm{K}_s$. Let us remark also that the image of a word not in F_s (resp. F_{ds}) by η_s (resp. η_{ds}) has an absorbing zero as image by η_s (resp. η_{ds}). This zero being idempotent, it belongs to $\mathcal{I}_E(S_s)$ (resp. $\mathcal{I}_E(S_{ds})$). Finally, if two words of F_s have the same image by η_s, then they have the same length modulo s and their first (and consequently last) letters have the same enrichment.

Consider then x a non-zero element of $\mathcal{I}_E(S_s)$. We show that

$$h^{-1}(\eta_s^{-1}(x)) \cap \eta_{ds}^{-1}(\mathcal{I}_E(S_{ds})) \neq \varnothing \ .$$

Since x belongs to $\mathcal{I}_E(S_s)$, there exists a word u of A_s^+ of length greater than s in the preimage of x. And since $\eta_s(A_s^s) = \eta_s(A_s^{2s})$ by definition of the stability index, for any $k > 0$ there exists a word v_k of A_s^+ of length greater than ks such that $u \equiv_L v_k$ and $|u| = |v_k| \bmod s$, since $\eta_s(u) = \eta_s(v_k)$. Then for k sufficiently large, there exists a word w in $h^{-1}(v_k)$, such that $\eta_{ds}(w)$ belongs to $\mathcal{I}_E(S_{ds})$. Note that by taking k as a multiple of d, we obtain a word w such that $|u| \bmod s = |w| \bmod ds$. Thus for each element $x \in \mathcal{I}_E(S_s)$, we can choose such an element, that we denote w_x. This justifies the definition of the following function:

$$f: \begin{cases} \mathcal{I}_E(S_s) & \to & \mathcal{I}_E(S_{ds}) & \\ x & \mapsto & \eta_{ds}(w_x) & \text{if } x \neq 0 \\ 0 & \mapsto & 0 & \text{otherwise.} \end{cases}$$

We conclude by proving that f is an injective morphism, and thus $\mathcal{I}_E(S_s)$ is a subsemigroup of $\mathcal{I}_E(S_{ds})$.

The application f is a morphism. Let $x, y \in \mathcal{I}_E(S_s)$. We show that $f(xy) = f(x)f(y)$. First, we can assume without loss of generality that $x \neq 0$ and $y \neq 0$. We remark that since $|w_x| \bmod ds = |h(w_x)| \bmod s$, the concatenated word $w_x w_y$ is well-formed if, and only if, $h(w_x)h(w_y)$ is well-formed too. If $xy \neq 0$.Then, xy have a well-formed preimage and $w_x w_y$ is well-formed. Then as w_{xy} and $w_x w_y$ are syntactically equivalent with respect to both F_{ds} and $h^{-1}(L_s)$, $\eta_{ds}(w_{xy}) = \eta_{ds}(w_x w_y) = \eta_{ds}(w_x)\eta_{ds}(w_y)$, meaning that $f(xy) = f(x)f(y)$. Now if $xy = 0$, then either xy has no well-formed preimage or xy is a zero for $\pi_s^{-1}(L)$. In the latter case, then $f(x)f(y) = 0$ according to the previous point. If xy has no well-formed preimage, then $w_x w_y$ is not well-formed and consequently $f(x)f(y) = 0$.

The application f is injective. Let $x, y \in \mathcal{I}_E(S_s)$ be such that $x \neq y$. Without loss of generality, we assume that $x \neq 0$. Necessarily, there exist $p, q \in S_s$ such that $pxq \in \eta_s(L_s)$ if, and only if, $pyq \notin \eta_s(L_s)$. Let u and v be words from the preimage of p and q respectively. Then there exists two words $u' \in h^{-1}(u) \cap \mathrm{F}_{ds}$ and $v' \in h^{-1}(v) \cap \mathrm{F}_{ds}$ such that $u' w_x v' \in L_{ds}$ if, and only if, $u' w_y v' \notin L_{ds}$. Therefore, we have $f(x) \neq f(y)$ and f is injective.

□

The following proposition deals with fragments which are not varieties of groups. Varieties of groups are notoriously ill behaving with respect to their global. Indeed Auinger (2010) exhibited a variety of group $\mathbf{H}$ such that $\mathbf{g}(\mathbf{LH})$ is undecidable (as a variety of *semigroupoids*). However, for a local variety $\mathbf{V}$ which is not a variety of groups, the variety of semigroups $\mathbf{LV}$ is local, as proved in Paperman (2014). Since this article does not deal with the framework of varieties of *semigroupoids*, we provide a self contain proof extracted from this latter result.

Proposition 29. *Let $\mathcal{F}[\sigma]$ be a fragment equivalent to a local variety $\mathbf{V}$ which is not a variety of groups. Then $\mathcal{F}[\sigma, +1, \mathrm{MOD}]$ is equivalent to $\mathbf{QLV}$.*

Proof: First we remark that since $\mathcal{F}[\sigma]$ is equivalent to a local variety, by Proposition 13, and by definition of locality, $\mathcal{F}[<, +1]$ is equivalent to $\mathbf{LV} = \mathbf{V} * \mathbf{D}$. Furthermore, since $\mathbf{V}$ is not a variety of groups, $(ab)^*$ belongs to $\mathbf{LV}$. Therefore, we obtain the following:

$$L \in \mathcal{F}[<, +1, \mathrm{MOD}] \underbrace{\text{ if and only if }}_{\text{By Theorem 28}} L \in \mathcal{F}[<, +1, \mathrm{MOD}^s] \underbrace{\text{ if and only if }}_{\text{By Theorem 3}} L_s \in \mathcal{F}[<, +1]$$

Claim: L_s belongs to **LV** if and only if L belongs to **QLV**.

We now prove both implications of the claim. In the following S_s will be the syntactic semigroup of L_s and S the one of L.

- Assume that L_s belongs to **LV**. Let $T = (A_s^s)^+ \cap \mathrm{K}_s$. We remark that T is a semigroup. Therefore, the set $\eta_s(T)$ is a subsemigroup of S_s. Since S_s belongs to **LV**, the semigroup $\eta_s(T)$ belongs to **LV** as well. Remark now that S_s is a quotient of the product of S and the syntactic semigroup of K_s. Since the image of $\pi_s(T)$ in the syntactic monoid of L is the stable semigroup of L and the image of T in the syntactic semigroup of K_s is trivial, we can conclude as $\eta_s(T)$ is isomorphic to the stable semigroup of L.

- Assume that L belongs to **QLV**, and we denote by T its stable semigroup. By hypothesis, T is in **LV**. One can remark that since **V** is not a variety of groups, it contains the semigroup $U_1 = \{0, 1\}$ (equipped with the integer multiplication). Therefore, the semigroup $T \cup \{0\}$, obtained by adding an absorbing element, also belongs to **LV**. Indeed, it divides $T \times U_1$.

  We now have to show that L_s is in **LV** as well. Let e be an idempotent of S_s. First, if e is the zero of S_s, then $eS_se = \{e\}$. Otherwise, e is the image of a well-formed factor u that starts by a letter of the form (a, i) and ends by a letter of the form (a, j) with $j + 1 \equiv i \bmod s$. We denote by f the image of $\pi_s(u)$ by the syntactic morphism of L. This element is idempotent and, therefore, belongs to T. We conclude by noting that the local monoid eS_se is a quotient of $fTf \cup \{0\}$.

□

6 Conclusion

In this paper, we studied the definability problem for fragments of logic enriched with the modular predicates. We presented a generic approach that gives the decidability of this problem in many cases, while the main applications are to the alternation hierarchies of the first order logic and its two variables counterpart.

The global approach is divided in two steps. The first one relies entirely on logic. We prove that adding a finite set of modular predicates preserves the decidability, given that the fragment is expressive enough.

The second part, which we call the delay problem, consists in deciding which finite set of modular predicates should be added to express a given regular language. This is the most intricate part of the paper. While unable to solve this question for any given fragment, we were able to reduce, following some known results, this question to a decidability question on the global of a fragment, a variety of categories. Then decidability was obtained for many fragments, using different approaches. They can be sorted in two cases. The first case is when the global is understood and finitely describable. Then we are able to decide a delay depending on the stability index and the said description. The second case is when the fragment is expressive enough to handle the modular predicates. This happens in particular if the fragment contains the local predicates and can use them extensively.

The main applications of these results are given in Figure 2, mainly on the levels of the quantifier alternation hierarchies, although this approach can be used on other fragments that satisfy the same hypotheses, such as the fragment **FO**[+1].

An interesting fact is that while the stability index often serves as a valid delay, this is still open whether this would hold for varieties of rank greater than two.

The question of solving the adding of modular predicate in a general setting seems achievable, although the more natural question would be to solve the decidability of the semidirect product by **MOD**. While we avoided this characterisation as it served no purpose in our approach, an aware reader could have noticed that Theorem 16 proves that adding modular predicates is algebraically equivalent to a semidirect product by the length-multiplying variety of morphisms **MOD**. Then our question reduces to whether this semidirect product preserves decidability. Auinger (2010) proved that the semidirect product in general does not preserve decidability, but the problem is still open for the case of **MOD**.

References

J. Almeida. A syntactical proof of locality of DA. *Internat. J. Algebra Comput.*, 6(2):165–177, 1996. ISSN 0218-1967. doi: 10.1142/S021819679600009X. URL `http://dx.doi.org/10.1142/S021819679600009X`.

K. Auinger. On the decidability of membership in the global of a monoid pseudovariety. *Internat. J. Algebra Comput.*, 20(2):181–188, 2010. ISSN 0218-1967. doi: 10.1142/S0218196710005571. URL `http://dx.doi.org/10.1142/S0218196710005571`.

D. A. M. Barrington, K. Compton, H. Straubing, and D. Thérien. Regular languages in NC^1. *J. Comput. System Sci.*, 44(3):478–499, 1992. ISSN 0022-0000. doi: 10.1016/0022-0000(92)90014-A. URL `http://dx.doi.org/10.1016/0022-0000(92)90014-A`.

J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik Grundlagen Math.*, 6:66–92, 1960.

L. Chaubard. *Méthodes algébriques pour les langages formels.* PhD dissertation, Université Paris Diderot, 2007.

L. Chaubard, J.-É. Pin, and H. Straubing. First order formulas with modular predicates. In *LICS*, pages 211–220. IEEE, 2006.

R. S. Cohen and J. A. Brzozowski. Dot-depth of star-free events. *J. Comput. System Sci.*, 5: 1–16, 1971. ISSN 0022-0000.

L. Dartois and C. Paperman. Two-variable first order logic with modular predicates over words. In N. Portier and T. Wilke, editors, *STACS*, LIPIcs, pages 329–340, Dagstuhl, Germany, 2013. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik.

L. Dartois and C. Paperman. Alternation hierarchies of first order logic with regular predicates. In A. Kosowski and I. Walukiewicz, editors, *Fundamentals of Computation Theory - 20th International Symposium, FCT 2015, Gdańsk, Poland, August 17-19, 2015, Proceedings*,

volume 9210 of *Lecture Notes in Computer Science*, pages 160–172. Springer, 2015. ISBN 978-3-319-22176-2. doi: 10.1007/978-3-319-22177-9_13. URL http://dx.doi.org/10.1007/978-3-319-22177-9_13.

V. Diekert, P. Gastin, and M. Kufleitner. A survey on small fragments of first-order logic over finite words. *Internat. J. Found. Comput. Sci.*, 19(3):513–548, 2008. ISSN 0129-0541. doi: 10.1142/S0129054108005802. URL http://dx.doi.org/10.1142/S0129054108005802.

Z. Ésik and M. Ito. Temporal logic with cyclic counting and the degree of aperiodicity of finite automata. *Acta Cybernet.*, 16(1):1–28, 2003. ISSN 0324-721X.

R. Knast. A semigroup characterization of dot-depth one languages. *RAIRO Inform. Théor.*, 17 (4):321–330, 1983. ISSN 0399-0540.

A. Krebs and H. Straubing. An effective characterization of the alternation hierarchy in two-variable logic. In *FSTTCS*, pages 86–98, 2012.

M. Kufleitner and A. Lauser. Lattices of logical fragments over words. In *ICALP (2)*, pages 275–286, 2012.

M. Kufleitner and A. Lauser. Quantifier alternation in two-variable first-order logic with successor is decidable. In *STACS*, pages 305–316, 2013.

M. Kufleitner and P. Weil. The FO^2 alternation hierarchy is decidable. In *Computer science logic 2012*, volume 16 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 426–439. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2012.

A. Maciel, P. Péladeau, and D. Thérien. Programs over semigroups of dot-depth one. *Theoret. Comput. Sci.*, 245(1):135–148, 2000. ISSN 0304-3975. doi: 10.1016/S0304-3975(99)00278-9. URL http://dx.doi.org/10.1016/S0304-3975(99)00278-9. Semigroups and algebraic engineering (Fukushima, 1997).

R. McNaughton and S. Papert. *Counter-free automata*. The M.I.T. Press, Cambridge, Mass.-London, 1971.

A. Nerode. Linear automaton transformation. In *Proceeding of the AMS*, volume 9, pages 541–544, 1958.

C. Paperman. *Circuits booléens, prédicats modulaires et langages réguliers*. PhD dissertation, Université Paris Diderot, 2014.

P. Péladeau. Logically defined subsets of $\mathbf{N}^k$. *Theoret. Comput. Sci.*, 93(2):169–183, 1992. ISSN 0304-3975. doi: 10.1016/0304-3975(92)90328-D. URL http://dx.doi.org/10.1016/0304-3975(92)90328-D.

J.-É. Pin. Syntactic semigroups. In *Handbook of formal languages, Vol. 1*, pages 679–746. Springer, Berlin, 1997.

J.-É. Pin. Profinite Methods in Automata Theory. In Susanne Albers and Jean-Yves Marion, editors, *26th International Symposium on Theoretical Aspects of Computer Science STACS 2009*, pages 31–50, Freiburg Allemagne, 2009. IBFI Schloss Dagstuhl. URL `http://hal.archives-ouvertes.fr/inria-00359677/en/`.

J.-É. Pin and H. Straubing. Some results on $\mathcal{C}$-varieties. *Theor. Inform. Appl.*, 39(1):239–262, 2005. ISSN 0988-3754. doi: 10.1051/ita:2005014. URL `http://dx.doi.org/10.1051/ita:2005014`.

T. Place and M. Zeitoun. Going higher in the first-order quantifier alternation hierarchy on words. In J. Esparza, P. Fraigniaud, T. Husfeldt, and E. Koutsoupias, editors, *ICALP, Proceedings, Part II*, volume 8573 of *Lecture Notes in Computer Science*, pages 342–353. Springer, 2014. ISBN 978-3-662-43950-0. doi: 10.1007/978-3-662-43951-7_29. URL `http://dx.doi.org/10.1007/978-3-662-43951-7`.

M. O. Rabin and D. Scott. Finite automata and their decision problems. *IBM Journal of Research and Development*, 3, 1959.

J. Rhodes and B. Steinberg. *The q-theory of finite semigroups*. Springer Monographs in Mathematics. Springer, New York, 2009. ISBN 978-0-387-09780-0. doi: 10.1007/b104443. URL `http://dx.doi.org/10.1007/b104443`.

M. P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965. ISSN 0890-5401.

I. Simon. Piecewise testable events. In *Automata theory and formal languages (Second GI Conf., Kaiserslautern, 1975)*, volume 33 of *Lect. Notes Comp. Sci.*, pages 214–222. Springer, Berlin, 1975.

H. Straubing. A generalization of the Schützenberger product of finite monoids. *Theoret. Comput. Sci.*, 13(2):137–150, 1981. ISSN 0304-3975. doi: 10.1016/0304-3975(81)90036-0. URL `http://dx.doi.org/10.1016/0304-3975(81)90036-0`.

H. Straubing. Finite semigroup varieties of the form $V * D$. *J. Pure Appl. Algebra*, 36(1):53–94, 1985. ISSN 0022-4049. doi: 10.1016/0022-4049(85)90062-3. URL `http://dx.doi.org/10.1016/0022-4049(85)90062-3`.

H. Straubing. *Finite automata, formal logic, and circuit complexity*. Birkhäuser Boston Inc., Boston, MA, 1994.

D. Thérien. Classification of finite monoids: the language approach. *Theoret. Comput. Sci.*, 14 (2 ang.):195–208, 1981. ISSN 0304-3975. doi: 10.1016/0304-3975(81)90057-8. URL `http://dx.doi.org/10.1016/0304-3975(81)90057-8`.

D. Thérien and T. Wilke. Over words, two variables are as powerful as one quantifier alternation. In *STOC '98 (Dallas, TX)*, pages 234–240. ACM, New York, 1999.

W. Thomas. Classifying regular events in symbolic logic. *J. Comput. System Sci.*, 25(3):360–376, 1982. ISSN 0022-0000. doi: 10.1016/0022-0000(82)90016-2. URL http://dx.doi.org/10.1016/0022-0000(82)90016-2.

B. Tilson. Categories as algebra: an essential ingredient in the theory of monoids. *J. Pure Appl. Algebra*, 48(1-2):83–198, 1987. ISSN 0022-4049. doi: 10.1016/0022-4049(87)90108-3. URL http://dx.doi.org/10.1016/0022-4049(87)90108-3.

P. Weis and N. Immerman. Structure theorem and strict alternation hierarchy for FO^2 on words. *Log. Methods Comput. Sci.*, 5(3):3:4, 23, 2009. ISSN 1860-5974. doi: 10.2168/LMCS-5(3:4)2009. URL http://dx.doi.org/10.2168/LMCS-5(3:4)2009.