

# All about unambiguous polynomial closure

Received Aug 17, 2022  
Revised Oct 30, 2022  
Accepted Nov 6, 2023  
Published Dec 23, 2023

## Key words and phrases

Words, regular languages, concatenation hierarchies, first-order logic, quantifier alternation, membership, separation

Thomas Place<sup>a</sup> ✉ 

Marc Zeitoun<sup>a</sup> ✉ 

<sup>a</sup> Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France

**ABSTRACT.** We study a standard operator on classes of languages: unambiguous polynomial closure. We prove that for every class  $C$  of regular languages satisfying mild properties, the membership problem for its unambiguous polynomial closure  $UPol(C)$  reduces to the same problem for  $C$ . We also show that unambiguous polynomial closure coincides with alternating left and right deterministic closure. Moreover, we prove that if additionally  $C$  is finite, the separation and covering problems are decidable for  $UPol(C)$ . Finally, we present an overview of the generic logical characterizations of the classes built using unambiguous polynomial closure.

## 1. Introduction

Regularity is arguably one of the most robust notions in computer science: indeed, regular languages can be equivalently defined using deterministic or nondeterministic automata, monoids, regular expressions or monadic second-order logic. The motivation and context of this paper is the investigation of *subclasses* of regular languages. Such classes arise naturally when weakening or restricting these formalisms. This active research track started in the 1960s with an emblematic example: a series of results by Schützenberger [34], McNaughton, Papert [14] and Kamp [11] established several equivalent characterizations of the class of languages defined by a *first-order* sentence (instead of a monadic second-order one). At that time, the main question that emerged was the *membership problem*, which simply asks whether the class under study is recursive: in order to solve it, one has to design an algorithm testing membership of an input language in the class. For instance, the above results provide a membership algorithm for first-order definable languages.

Work supported by the DeLTA project (ANR-16-CE40-0007), and, for the first author, by the Institut Universitaire de France.

Of course, the design of a membership algorithm for a class  $C$  heavily depends on  $C$ . This means that for each class, one has to design a different algorithm. However, most of the interesting classes of regular languages are built using a restricted set of operators: given a class  $C$ , one can consider its Boolean closure  $Bool(C)$ , its polynomial closure  $Pol(C)$ , and deterministic variants thereof, which usually yield more elaborate classes than  $C$ . It is therefore desirable to investigate the operators themselves rather than individual classes.

The *polynomial closure*  $Pol(C)$  of a class  $C$  is its closure under finite union and marked concatenation (a *marked concatenation* of  $K$  and  $L$  is a language of the form  $KaL$ , where  $a$  is some letter). Together with the Boolean closure, it is used to define concatenation hierarchies, which consist in a sequence of classes of languages indexed by integers and half integers. Starting from a given class (*level 0* in the hierarchy), level  $n + \frac{1}{2}$  is the polynomial closure of level  $n$ , and level  $n + 1$  is the Boolean closure of level  $n + \frac{1}{2}$ . The importance of these hierarchies stems from the fact that they are the combinatorial counterpart of quantifier alternation hierarchies in logic, which count the number of  $\forall/\exists$  alternations needed to define a language [41, 26].

As explained above, the main question when investigating a class of languages is the *membership* problem: can we decide whether an input regular language belongs to the class? Despite decades of research on concatenation hierarchies, we know little about them. The state of the art is that when level 0 is either *finite* or contains only *group languages* (and satisfies mild properties), membership is decidable for levels  $1/2$ , one,  $3/2$ , and  $5/2$  [31, 22, 27, 30] (on the other hand, the problem remains open for the second full level). These results encompass those that were obtained previously [3, 4, 36, 20] and even go beyond by investigating the *separation problem*, a generalization of membership. This problem for a class  $C$  takes *two arbitrary regular* languages as input (unlike membership, which takes a single one). It asks whether there exists a third language from  $C$ , containing the first language and disjoint from the second. Membership is the special case of separation when the input consists of a language and its complement (as the only possible separator is the first input language). Although more difficult than membership, separation is also more rewarding. This is witnessed by a transfer theorem [26]: membership for  $Pol(C)$  reduces to separation for  $C$ . The above results on membership come from this theorem and the fact that separation is decidable for  $Pol(C)$ ,  $BPol(C)$  (i.e.,  $Bool(Pol(C))$ ) and  $Pol(BPol(C))$  when  $C$  is either *finite* or contains only *group languages*. See [18, 26] for detailed surveys on concatenation hierarchies.

**Unambiguous closure.** A natural way to weaken the polynomial closure operator is to specify *semantic conditions* restricting the situations in which using marked concatenation or union is allowed. In the paper, we investigate a prominent operator that can be defined in this way: *unambiguous polynomial closure*. A marked concatenation  $KaL$  is *unambiguous* if every word  $w$  belonging to  $KaL$  has a *unique* factorization  $w = uav$  with  $u \in K$  and  $v \in L$ . The *unambiguous closure* of a class  $C$ , denoted by  $UPol(C)$ , is the least class containing  $C$  that is closed under

*disjoint* union and *unambiguous* marked concatenation. Observe that it is not immediate from the definition that  $UPol(C)$  has robust properties (such as closure under Boolean operations) even when  $C$  does. A prominent example of a class built using unambiguous concatenation is that of *unambiguous languages* [35] (UL). It is the unambiguous polynomial closure of the class  $BPol(ST)$  (i.e.,  $UPol(BPol(ST))$ ) where  $ST$  is the trivial class  $ST = \{\emptyset, A^*\}$ <sup>1</sup>. The robustness of UL makes it one of the most investigated classes: it enjoys a number of equivalent definitions (we refer the reader to [38, 8] for an overview). In particular, this class has several logical characterizations. First, UL corresponds to the intermediary level  $\Delta_2(<)$  in the quantifier alternation hierarchy of first-order logic equipped with the linear order predicate [20]. This level consists of all languages that can be simultaneously defined by a sentence of  $\Sigma_2(<)$  (i.e., whose prenex normal form has a quantifier prefix of the form  $\exists^*\forall^*$ ) and a sentence of  $\Pi_2(<)$  (i.e., whose prenex normal form has a quantifier prefix of the form  $\forall^*\exists^*$ ). It is also known that UL corresponds to the two-variable fragment  $FO^2(<)$  of first-order logic [39]. Finally, UL is also characterized [9] by the variant  $F + P$  of *unary temporal logic*, where  $F$  stands for “sometimes in the future” and  $P$  stands for “sometimes in the past”.

Historically, the operator  $UPol$  was first investigated by Pin [15] who describes it in algebraic terms. Note however that [15] starts from an alternative definition that assumes closure under Boolean operations already. It was later shown by Pin, Straubing and Thérien [19] that the operator  $C \mapsto UPol(C)$  preserves closure under Boolean operations (provided that the input class  $C$  satisfies mild hypotheses). Let us point out that both the formulation and the proof of these results rely on elaborate mathematical tools (categories, bilateral kernel, relational morphisms,...) as well as on results by Schützenberger [35] and Rhodes [33], used as black boxes. Unambiguous polynomial closure also appears in concatenation hierarchies: assuming suitable hypotheses on  $C$ , Pin and Weil [20] proved that  $UPol(C)$  is the *intermediate level*  $Pol(C) \cap coPol(C)$  where  $coPol(C)$  is the class consisting of all complements of languages in  $Pol(C)$ . Finally, a reduction from  $UPol(C)$ -membership to  $C$ -membership was obtained in [2]. This proof is indirect: it relies on the nontrivial equality  $UPol(C) = Pol(C) \cap coPol(C)$ , which itself depends on the algebraic characterizations of  $UPol(C)$  and  $Pol(C)$  obtained in [19, 20, 6].

**Contributions.** Unambiguous polynomial closure was not yet investigated with respect to separation, aside from the particular case of unambiguous languages [24]. This is the starting point of this paper: we look for generic separation results applying to  $UPol(C)$ , similar to the ones obtained for  $Pol(C)$  in [22]. We prove that when  $C$  is a finite Boolean algebra closed under quotients, separation is decidable for  $UPol(C)$ . However, as is usual with separation, we also obtain several extra results as a byproduct of our work, improving our understanding of the  $UPol$  operator:

---

1 The notation  $ST$  comes from the fact that it is level 0 in the Straubing-Thérien concatenation hierarchy.

- We had to rethink the way membership is classically handled for  $UPol$  in order to lift the techniques to separation. This yields self-contained, direct and elementary proofs that membership for  $UPol(C)$  reduces to membership for  $C$  provided that  $C$  is closed under Boolean operations and quotients. The proof is new and independent from the one of [2]. We strengthen the result of [2] as we require fewer hypotheses on the class  $C$ . Our proof also precisely pinpoints why this result holds for  $UPol(C)$  but not  $Pol(C)$ : the languages from  $C$  needed to construct a  $UPol(C)$  expression for a language  $L$  are all recognized by any recognizer of  $L$ .
- We prove that when  $C$  is closed under Boolean operations and quotients, then so is  $UPol(C)$ . Again, this strengthens earlier results [19], which involved stronger hypotheses on the input class  $C$ .
- Using our results on  $Pol(C)$  [26], we obtain a new proof of the above mentioned theorem of [20] and generalize it to more classes  $C$ . We prove that  $UPol(C) = Pol(C) \cap coPol(C)$  when  $C$  is closed under Boolean operations and quotients (whereas the result of [20] only applies to classes that are additionally closed under inverse morphic images).
- Finally, we obtain a previously unknown characterization of  $UPol(C)$  in terms of alternating left and right deterministic concatenations, which are restricted forms of unambiguous concatenation. A marked concatenation  $KaL$  is *left (resp. right) deterministic* when  $KaA^* \cap K = \emptyset$  (resp.  $A^*aL \cap L = \emptyset$ ). We prove that  $UPol(C)$  coincides with  $APol(C)$ , the closure of  $C$  under disjoint union and left and right deterministic concatenation<sup>2</sup>.

Moreover, we investigate the logical characterizations of unambiguous polynomial closure. As mentioned above, it is well known that  $UPol(BPol(ST)) = \Delta_2(<) = FO^2(<) = F + P$ . Such correspondences also hold for variants of  $\Delta_2(<)$  and  $FO^2(<)$  equipped with additional predicates such as the successor [39] “+1” and the modular predicates [7, 13]. We use our results to generalize these connections. For each class  $C$ , we define two objects:

- A set of predicates  $\mathbb{P}_C$  that can be used in the sentences of first-order logic. Each language  $L$  in  $C$  gives rise to a predicate  $P_L(x)$ , which selects all positions  $x$  in a word  $w$  such that the prefix of  $w$  up to position  $x$  (excluded) belongs to  $L$ .
- Two variants of unary temporal logic, which we denote by  $TL(C)$  and  $TLX(C)$ .

We consider classes of group languages. *Group languages* are those recognized by a finite group, or equivalently by a permutation automaton [40] (i.e., a complete, deterministic and co-deterministic automaton). Our results apply to classes of group languages  $\mathcal{G}$  closed under Boolean operations and quotients, and their well-suited extensions  $\mathcal{G}^+$  (roughly,  $\mathcal{G}^+$  is the

---

2 The letter “A” in  $APol(C)$  stands for “alternating” (left and right) deterministic polynomial closure.

smallest Boolean algebra containing  $\mathcal{G}$  and the singleton  $\{\varepsilon\}$ ). In such cases, we prove that,

$$\begin{aligned} UPol(BPol(\mathcal{G})) &= \Delta_2(<, \mathbb{P}_{\mathcal{G}}) = FO^2(<, \mathbb{P}_{\mathcal{G}}) = TL(\mathcal{G}). \\ UPol(BPol(\mathcal{G}^+)) &= \Delta_2(<, +1, \mathbb{P}_{\mathcal{G}}) = FO^2(<, +1, \mathbb{P}_{\mathcal{G}}) = TLX(\mathcal{G}). \end{aligned}$$

This generalizes all aforementioned results (each of them corresponds to a particular class of group languages  $\mathcal{G}$ ). Thus, we obtain a generic proof of these results.

**Organization of the paper.** Section 2 sets up the notation and the terminology. In Section 3, we recall the definition of first-order logic over words and introduce the fragments that we shall consider. Section 4 is devoted to presenting the operators that we investigate in the paper. In Section 5, we introduce notions that we shall need to present the algebraic characterizations of  $Pol$  and  $UPol$ . Section 6 presents the algebraic characterization of  $Pol(C)$  (which is taken from [26]). We use it to obtain a characterization of the classes  $Pol(C) \cap coPol(C)$ . Then, in Section 7, we turn to the classes  $UPol(C)$  and present their generic characterization. We investigate separation and covering for the classes  $UPol(C)$  when  $C$  is finite in Section 8. Finally, Sections 9 and 10 are devoted to the logical characterizations of  $UPol$ . In the former, we define our generalized notion of “unary temporal logic” and prove the correspondence with two-variable first-order logic. Finally, we establish the connection with unambiguous polynomial closure in Section 10.

This paper is the journal version of [29]. The presentation and the proof arguments have been reworked. Moreover, the logical part of the paper is new (the conference version only considered the language theoretic point of view).

## 2. Preliminaries

We introduce the terminology used in the paper. In particular, we present the membership, separation and covering problems, and key mathematical tools designed to handle them.

### 2.1 Classes of regular languages

For the whole paper, we fix a finite alphabet  $A$ . As usual,  $A^*$  denotes the set of all *finite* words over  $A$ , including the empty word  $\varepsilon$ . Moreover, we let  $A^+ = A^* \setminus \{\varepsilon\}$ . Given a word  $w \in A^*$ , we write  $|w| \in \mathbb{N}$  for its length (the number of letters in  $w$ ). For example,  $|\varepsilon| = 0$  and  $|abcbaa| = 6$ . Moreover, for every  $a \in A$  and  $w \in A^*$ , we write  $|w|_a \in \mathbb{N}$  for the number of copies of the letter “ $a$ ” in  $w$ . For example,  $|abcbaa|_a = 3$ ,  $|abcbaa|_b = 2$  and  $|abcbaa|_c = 1$ . Given  $u, v \in A^*$ , we write  $uv$  for their concatenation.

A *language* is a subset of  $A^*$ . It is standard to extend concatenation to languages: given  $K, L \subseteq A^*$ , we write  $KL = \{uv \mid u \in K \text{ and } v \in L\}$ .

**Classes.** A class of languages  $C$  is a set of languages. Such a class  $C$  is a *lattice* when  $\emptyset \in C$ ,  $A^* \in C$  and  $C$  is closed under both union and intersection: for every  $K, L \in C$ , we have  $K \cup L \in C$  and  $K \cap L \in C$ . Moreover, a *Boolean algebra* is a lattice  $C$  which is additionally closed under complement: for every  $L \in C$ , we have  $A^* \setminus L \in C$ . Finally, a class  $C$  is *quotient-closed* if for every  $L \in C$  and  $u \in A^*$ , the following properties hold:

$$u^{-1}L \stackrel{\text{def}}{=} \{w \in A^* \mid uw \in L\} \quad \text{and} \quad Lu^{-1} \stackrel{\text{def}}{=} \{w \in A^* \mid wu \in L\} \quad \text{both belong to } C.$$

We call a positive prevariety (resp. prevariety) a quotient-closed lattice (resp. Boolean algebra) containing only *regular languages*. Regular languages are those which can be equivalently defined by nondeterministic finite automata, finite monoids or monadic second-order logic. We work with the definition by monoids, which we recall below.

In the paper, we write AT for the class of *alphabet testable languages*: the Boolean combinations of languages  $A^*aA^*$  for  $a \in A$ . It is straightforward to verify that AT is a prevariety. It will be important for providing examples.

**Finite monoids and morphisms.** A *semigroup* is a set  $S$  endowed with an associative multiplication  $(s, t) \mapsto s \cdot t$  (also denoted by  $st$ ). A *monoid* is a semigroup  $M$  whose multiplication has an identity element  $1_M$ , i.e., such that  $1_M \cdot s = s \cdot 1_M = s$  for every  $s \in M$ .

An *idempotent* of a semigroup  $S$  is an element  $e \in S$  such that  $ee = e$ . We write  $E(S) \subseteq S$  for the set of all idempotents in  $S$ . It is folklore that for every *finite* semigroup  $S$ , there exists a natural number  $\omega(S)$  (denoted by  $\omega$  when  $S$  is understood) such that for every  $s \in S$ , the element  $s^\omega$  is an idempotent.

We also consider ordered semigroups. An *ordered semigroup* is a pair  $(S, \leq)$  where  $S$  is a semigroup and  $\leq$  is a partial order defined on  $S$ , which is compatible with multiplication: for every  $s, s', t, t' \in S$  such that  $s \leq t$  and  $s' \leq t'$ , we have  $ss' \leq tt'$ . In particular, we say that a subset  $F \subseteq S$  is an *upper set* (for  $\leq$ ) to indicate that it is upward closed for the ordering  $\leq$ : for every  $s, t \in M$ , if  $s \in F$  and  $s \leq t$ , then  $t \in F$ . In particular, for every  $s \in S$ , we write  $\uparrow s \subseteq S$  for the least upper set containing  $s$ . That is,  $\uparrow s = \{t \in S \mid s \leq t\}$ . Finally, an *ordered monoid* is an ordered semigroup  $(M, \leq)$  such that  $M$  is a monoid.

We shall use the following convention throughout the paper: we view every *unordered* semigroup  $S$  as the ordered semigroup  $(S, =)$  whose ordering is *equality*. Observe that in this case, *every* subset of  $S$  is an upper set. Thanks to this convention, every definition involving ordered semigroups that we present will also make sense for unordered ones.

Finally, our proofs make use of the Green relations [10], which are defined on monoids. We briefly recall them. Given a monoid  $M$  and  $s, t \in M$ ,

$$\begin{aligned} s \leq_{\mathcal{J}} t & \text{ when there exist } x, y \in M \text{ such that } s = xty, \\ s \leq_{\mathcal{L}} t & \text{ when there exists } x \in M \text{ such that } s = xt, \\ s \leq_{\mathcal{R}} t & \text{ when there exists } y \in M \text{ such that } s = ty. \end{aligned}$$

Clearly,  $\leq_{\mathcal{J}}$ ,  $\leq_{\mathcal{L}}$  and  $\leq_{\mathcal{R}}$  are preorders (i.e., they are reflexive and transitive). We write  $<_{\mathcal{J}}$ ,  $<_{\mathcal{L}}$  and  $<_{\mathcal{R}}$  for their strict variants (for example,  $s <_{\mathcal{J}} t$  when  $s \leq_{\mathcal{J}} t$  but  $t \not\leq_{\mathcal{J}} s$ ). Finally, we write  $\mathcal{J}$ ,  $\mathcal{L}$  and  $\mathcal{R}$  for the corresponding equivalence relations (for example,  $s \mathcal{J} t$  when  $s \leq_{\mathcal{J}} t$  and  $t \leq_{\mathcal{J}} s$ ). There are many technical results about Green relations. We use the following standard lemma (see e.g. [17] for a proof), which applies to *finite* monoids.

**LEMMA 2.1.** Consider a finite monoid  $M$  and  $s, t \in M$  such that  $s \mathcal{J} t$ . Then,  $s \leq_{\mathcal{R}} t$  implies  $s \mathcal{R} t$ . Symmetrically,  $s \leq_{\mathcal{L}} t$  implies  $s \mathcal{L} t$ .

**Regular languages and syntactic morphisms.** Clearly,  $A^*$  is a monoid whose multiplication is concatenation (the identity element is the empty word  $\varepsilon$ ). We shall consider monoid morphisms  $\alpha : A^* \rightarrow (M, \leq)$  where  $(M, \leq)$  is an arbitrary ordered monoid. That is,  $\alpha : A^* \rightarrow M$  is a map satisfying  $\alpha(\varepsilon) = 1_M$  and  $\alpha(uv) = \alpha(u)\alpha(v)$  for all  $u, v \in A^*$ . Given such a morphism and some language  $L \subseteq A^*$ , we say that  $L$  is *recognized* by  $\alpha$  when there exists an accepting set  $F \subseteq M$  which is an upper set for  $\leq$  and such that  $L = \alpha^{-1}(F)$ . Let us emphasize that the definition depends on the ordering  $\leq$  since accepting sets must be upper sets.

**REMARK 2.2.** Recall that by convention, we view an unordered monoid  $M$  as the ordered monoid  $(M, =)$ . Hence, the definition also makes sense when  $\alpha : A^* \rightarrow M$  is a morphism into an unordered monoid. Since every subset of  $M$  is an upper set for  $=$ , a language  $L \subseteq A^*$  is recognized by  $\alpha$  if and only if there exists an arbitrary set  $F \subseteq M$  such that  $L = \alpha^{-1}(F)$ .

It is standard and well known that regular languages are those which can be recognized by a morphism into a *finite* monoid. Moreover, every language  $L$  is recognized by a canonical morphism. Let us briefly recall its definition. One may associate to  $L$  a preorder relation  $\leq_L$  over  $L$ : the *syntactic precongruence* of  $L$ . Given  $u, v \in A^*$ , we let,

$$u \leq_L v \text{ if and only if } xuy \in L \Rightarrow xvy \in L \text{ for every } x, y \in A^*.$$

As the name suggests, it is known and simple to verify that “ $\leq_L$ ” is a precongruence on  $A^*$ : it is reflexive and transitive, and for every  $u, u', v, v' \in A^*$  such that  $u \leq_L v$  and  $u' \leq_L v'$ , we have  $uu' \leq_L vv'$ . Additionally, we write  $\equiv_L$  for the equivalence generated by this preorder. For  $u, v \in A^*$ ,  $u \equiv_L v$  if and only if  $u \leq_L v$  and  $v \leq_L u$  (i.e.,  $xuy \in L \Leftrightarrow xvy \in L$  for every  $x, y \in A^*$ ). The equivalence  $\equiv_L$  is a congruence called the *syntactic congruence* of  $L$ . Thus, the set of equivalence classes  $M_L = A^*/\equiv_L$  is a monoid and the preorder  $\leq_L$  defines an ordering  $\leq_L$  on  $M_L$  such that  $(M_L, \leq_L)$  is an ordered monoid. It is called the *syntactic ordered monoid* of  $L$ . Finally, the map  $\alpha_L : A^* \rightarrow (M_L, \leq_L)$  sending every word to its equivalence class is a morphism recognizing  $L$ , called the *syntactic morphism* of  $L$ . Another characterization of the regular languages is that  $L$  is regular if and only if  $M_L$  is finite (i.e.,  $\equiv_L$  has finite index): this is the Myhill-Nerode theorem. In this case, one can compute the syntactic morphism  $\alpha_L : A^* \rightarrow (M_L, \leq_L)$  from any representation of  $L$  (such as a finite automaton or an arbitrary monoid morphism).

## 2.2 Decision problems

We consider three decision problems. They all depend on an arbitrary class of languages  $C$ : they serve as mathematical tools for analyzing  $C$ . Intuitively, obtaining an algorithm for one of these three problems requires a solid understanding of  $C$ .

The *C-membership problem* is the simplest one. It takes as input a single regular language  $L$  and asks whether  $L \in C$ . The second problem, *C-separation*, is more general. Given three languages  $K, L_1, L_2$ , we say that  $K$  separates  $L_1$  from  $L_2$  if we have  $L_1 \subseteq K$  and  $L_2 \cap K = \emptyset$ . Given a class of languages  $C$ , we say that  $L_1$  is *C-separable* from  $L_2$  if some language in  $C$  separates  $L_1$  from  $L_2$ . Observe that when  $C$  is not closed under complement, the definition is not symmetrical: it is possible for  $L_1$  to be *C-separable* from  $L_2$  while  $L_2$  is not *C-separable* from  $L_1$ . The separation problem associated to a given class  $C$  takes two regular languages  $L_1$  and  $L_2$  as input and asks whether  $L_1$  is *C-separable* from  $L_2$ .

**REMARK 2.3.** The *C-separation* problem generalizes the *C-membership* problem. Indeed, a regular language belongs to  $C$  if and only if it is *C-separable* from its complement, which is also regular.

In the paper, we do not consider separation directly. Instead, we work with a third, even more general problem: *C-covering*. It was introduced in [32] and takes as input a single regular language  $L_1$  and a finite set of regular languages  $L_2$ . It asks whether there exists a “*C-cover* of  $L_1$  which is separating for  $L_2$ ”.

Given a language  $L$ , a *cover* of  $L$  is a finite set of languages  $\mathbf{K}$  such that  $L \subseteq \bigcup_{K \in \mathbf{K}} K$ . A cover  $\mathbf{K}$  is a *C-cover* if all languages  $K \in \mathbf{K}$  belong to  $C$ . Moreover, given two finite sets of languages  $\mathbf{K}$  and  $\mathbf{L}$ , we say that  $\mathbf{K}$  is *separating* for  $\mathbf{L}$  if for every  $K \in \mathbf{K}$ , there exists  $L \in \mathbf{L}$  such that  $K \cap L = \emptyset$ . Finally, given a language  $L_1$  and a finite set of languages  $L_2$ , we say that the pair  $(L_1, L_2)$  is *C-coverable* if there exists a *C-cover* of  $L_1$  which is separating for  $L_2$ .

The *C-covering* problem is now defined as follows. Given as input a regular language  $L_1$  and a finite set of regular languages  $L_2$ , it asks whether the pair  $(L_1, L_2)$  is *C-coverable*. It is straightforward to prove that covering generalizes separation if the class  $C$  is a lattice, as stated in the following lemma (see [32, Theorem 3.5] for the proof, which is easy).

**LEMMA 2.4.** Let  $C$  be a lattice and  $L_1, L_2$  be two languages. Then  $L_1$  is *C-separable* from  $L_2$  if and only if  $(L_1, \{L_2\})$  is *C-coverable*.

## 2.3 Group languages

We now present an important special kind of class. As we explained in the introduction, all classes investigated in the paper are built from basic ones using generic operators. Here, we introduce the basic classes used in such constructions: the *classes of group languages* and their *well-suited extensions*.

**DEFINITION 2.5.** A *group* is a monoid  $G$  such that every element  $g \in G$  has an inverse  $g^{-1} \in G$ , i.e.,  $gg^{-1} = g^{-1}g = 1_G$ . A language  $L$  is a *group language* if it is recognized by a morphism  $\alpha : A^* \rightarrow G$  into a *finite group*  $G$  (i.e., there exists  $F \subseteq G$  such that  $L = \alpha^{-1}(F)$ ). We write GR for the class of *all* group languages. One can verify that GR is a prevariety.

**REMARK 2.6.** No language theoretic definition of GR is known. There is however a definition based on automata: the group languages are those which can be recognized by a permutation automaton (i.e., which is simultaneously deterministic, co-deterministic and complete).

A *class of group languages* is a class consisting of group languages only, i.e., a subclass of GR. The results of this paper apply to arbitrary *prevarieties of group languages*.

**REMARK 2.7.** Let us explain why we do not explicitly mention *positive prevarieties* of group languages in the paper. This is because they are all closed under complement (hence, they are actually *prevarieties*). This follows from the simple fact that the only ordering compatible with multiplication in a finite group is the equality relation.

While our results apply in a generic way to all prevarieties of group languages, there are *four* main classes of this kind that we shall use for providing examples. One of them is GR itself. Let us present the other three. First, we write  $ST = \{\emptyset, A^*\}$ , which is clearly a prevariety of group languages (the name comes from the fact that this class is the base level of the *Straubing-Thérien* hierarchy). While trivial, we shall see that this class has important applications. Moreover, we look at the class MOD of *modulo languages*. For every  $q, r \in \mathbb{N}$  with  $r < q$ , we write  $L_{q,r} = \{w \in A^* \mid |w| \equiv r \pmod{q}\}$ . The class MOD contains all *finite unions* of languages  $L_{q,r}$ . One can verify that MOD is a prevariety of group languages. Finally, we shall consider the class AMT of *alphabet modulo testable languages*. For all  $q, r \in \mathbb{N}$  with  $r < q$  and all  $a \in A$ , let  $L_{q,r}^a = \{w \in A^* \mid |w|_a \equiv r \pmod{q}\}$ . We define AMT as the least class consisting of all languages  $L_{q,r}^a$  and closed under union and intersection. It is again straightforward to verify that AMT is a prevariety of group languages.

We do not investigate classes of group languages themselves in the paper: we only use them as input classes for our operators. In particular, we shall use ST, MOD, AMT and GR in order to illustrate our results. In this context, it will be important that *separation* is decidable for these four classes. The techniques involved in proving this are independent of what we do in the paper. Actually, this can be difficult. On one hand, the decidability of ST-separation is immediate (two languages are ST-separable if and only if one of them is empty). On the other hand, the decidability of GR-separation is equivalent to a difficult algebraic question, which remained open for several years before it was solved by Ash [5]. Recent automata-based proofs that separation is decidable for MOD, AMT and GR are available in [28].

**Well-suited extensions.** It can be verified from the definition that  $\{\varepsilon\}$  and  $A^+$  are *not* group languages. This motivates the next definition: for a class  $C$ , the *well-suited extension* of  $C$ ,

denoted by  $C^+$ , consists of all languages of the form  $L \cap A^+$  and  $L \cup \{\varepsilon\}$  where  $L \in C$  (in particular,  $C \subseteq C^+$ ). The next lemma follows immediately from the definition.

**LEMMA 2.8.** Let  $C$  be a prevariety. Then,  $C^+$  is a prevariety containing  $\{\varepsilon\}$  and  $A^+$ .

While the definition of well-suited extensions makes sense for arbitrary classes, it will only be useful when applied to a prevariety of group languages. Indeed, the well-suited extensions  $\mathcal{G}^+$ , where  $\mathcal{G}$  is a prevariety of group languages, are also important input classes for our operators. In this case, all group languages in  $\mathcal{G}^+$  already belong to  $\mathcal{G}$ .

**LEMMA 2.9.** Let  $\mathcal{G}$  be a prevariety of group languages. For every group language  $L \in \mathcal{G}^+$ , we have  $L \in \mathcal{G}$ .

**PROOF.** By definition of  $\mathcal{G}^+$ , there exists a group language  $K \in \mathcal{G}$  such that either  $L = \{\varepsilon\} \cup K$  or  $L = A^+ \cap K$ . In particular, this implies that for every  $w \in A^+$ , we have  $w \in K \Leftrightarrow w \in L$ . We prove that the fact that  $K$  and  $L$  are both group languages implies  $K = L$ , which yields  $L \in \mathcal{G}$ , as desired. Since we already know that  $w \in K \Leftrightarrow w \in L$  for every  $w \in A^+$ , it suffices to prove that  $\varepsilon \in K \Leftrightarrow \varepsilon \in L$ . Since  $K$  and  $L$  are group languages, there exist two morphisms  $\alpha_K : A^* \rightarrow G_K$  and  $\alpha_L : A^* \rightarrow G_L$  into finite groups recognizing  $K$  and  $L$  respectively. Let  $p = \omega(G_K) \times \omega(G_L)$  and consider an arbitrary letter  $a \in A$ . By definition of  $p$ , we have  $\alpha_K(a^p) = 1_{G_K}$  and  $\alpha_L(a^p) = 1_{G_L}$ . Since  $\alpha_K$  and  $\alpha_L$  recognize both  $K$  and  $L$ , we have  $\varepsilon \in K \Leftrightarrow a^p \in K$  and  $a^p \in L \Leftrightarrow \varepsilon \in L$ . Finally, since  $a^p \in A^+$ , we know that  $a^p \in K \Leftrightarrow a^p \in L$  by definition. Altogether, it follows that  $\varepsilon \in K \Leftrightarrow \varepsilon \in L$ , as desired. ■

## 2.4 $C$ -morphisms

We now present a central mathematical tool. Consider an arbitrary positive prevariety  $C$ . A  $C$ -morphism is a *surjective* morphism  $\eta : A^* \rightarrow (N, \leq)$  into a finite ordered monoid  $(N, \leq)$  such that every language recognized by  $\eta$  belongs to  $C$ . We complete the definition with a key remark when  $C$  is a prevariety (*i.e.*,  $C$  is additionally closed under complement): in this case, it suffices to consider *unordered* monoids (recall that by convention, we view them as monoids ordered by equality).

**LEMMA 2.10.** Let  $C$  be a prevariety and let  $\eta : A^* \rightarrow (N, \leq)$  be a morphism into a finite ordered monoid. The two following conditions are equivalent:

1.  $\eta : A^* \rightarrow (N, \leq)$  is a  $C$ -morphism.
2.  $\eta : A^* \rightarrow (N, =)$  is a  $C$ -morphism.

**PROOF.** The implication 2)  $\Rightarrow$  1) is trivial since all languages recognized by  $\eta : A^* \rightarrow (N, \leq)$  are also recognized by  $\eta : A^* \rightarrow (N, =)$ . We prove that 1)  $\Rightarrow$  2). Assume that  $\eta : A^* \rightarrow (N, \leq)$  is a  $C$ -morphism. We have to prove that  $\eta : A^* \rightarrow (N, =)$  is a  $C$ -morphism, that is, that for

all  $F \subseteq N$ , we have  $\eta^{-1}(F) \in C$ . Since  $\eta^{-1}(F) = \bigcup_{s \in F} \eta^{-1}(s)$  and  $C$  is closed under union, it suffices to prove that  $\eta^{-1}(s) \in C$  for all  $s \in N$ . Therefore, we fix an arbitrary element  $s \in N$ . Let  $T \subseteq M$  be the set of all elements  $t \in M$  such that  $s \leq t$  and  $s \neq t$ . One can verify that we have  $\{s\} = (\uparrow s) \setminus (\bigcup_{t \in T} \uparrow t)$ . Consequently, we obtain  $\eta^{-1}(s) = \eta^{-1}(\uparrow s) \setminus (\bigcup_{t \in T} \eta^{-1}(\uparrow t))$ . By hypothesis, this implies that  $\eta^{-1}(s)$  is a Boolean combination of languages in  $C$ . We conclude that  $\eta^{-1}(s) \in C$ , since  $C$  is a Boolean algebra. ■

**REMARK 2.11.** Another formulation of Lemma 2.10 is that when  $C$  is a prevariety, whether a morphism  $\eta : A^* \rightarrow (N, \leq)$  is a  $C$ -morphism does not depend on the ordering  $\leq$  on  $N$ . This is important, as most classes that we consider in the paper are indeed prevarieties.

While simple, the notion of  $C$ -morphism is a central tool in the paper. First, it is connected to the membership problem via the following simple, yet crucial proposition.

**PROPOSITION 2.12.** Let  $C$  be a positive prevariety. A regular language  $L$  belongs to  $C$  if and only if its syntactic morphism  $\alpha_L : A^* \rightarrow (M_L, \leq_L)$  is a  $C$ -morphism.

**PROOF.** The “if” implication is immediate since  $L$  is recognized by its syntactic morphism. We prove the converse direction: assuming that  $L \in C$ , we prove that every language recognized by  $\alpha_L : A^* \rightarrow (M_L, \leq_L)$  belongs to  $C$  (recall that syntactic morphisms are surjective by definition). Hence, we fix an upper set  $F \subseteq M_L$  and prove that  $\alpha_L^{-1}(F) \in C$ . For every  $s \in M_L$ , we fix a word  $u_s \in A^*$  such that  $\alpha_L(u_s) = s$ . Moreover, we let  $P_s \subseteq M_L^2$  be the set,

$$P_s = \{(q, r) \in M_L^2 \mid qsr \in \alpha_L(L)\}.$$

We now prove that,

$$\alpha_L^{-1}(F) = \bigcup_{s \in F} \bigcap_{(q,r) \in P_s} (u_q^{-1} L u_r^{-1}).$$

Since  $L \in C$  and  $C$  is a positive prevariety, this yields  $\alpha_L^{-1}(F) \in C$ , as desired. We start with the left to right inclusion. Consider  $w \in \alpha_L^{-1}(F)$  and let  $s = \alpha_L(w) \in F$ . We prove that  $w \in u_q^{-1} L u_r^{-1}$  for every  $(q, r) \in P_s$ . Clearly,  $\alpha_L(u_q w u_r) = qsr$  and by definition of  $P_s$ , we have  $qsr \in \alpha_L(L)$ . Since  $L$  is recognized by  $\alpha_L$ , we get  $u_q w u_r \in L$ , which exactly says that  $w \in u_q^{-1} L u_r^{-1}$ , as desired.

We turn to the converse inclusion. Fix  $s \in F$  and consider  $w \in \bigcap_{(q,r) \in P_s} (u_q^{-1} L u_r^{-1})$ . We show that  $w \in \alpha_L^{-1}(F)$ , i.e.,  $\alpha_L(w) \in F$ . Since  $s \in F$  and  $F$  is an upper set, it suffices to prove that  $s \leq_L \alpha_L(w)$ . Recall that  $u_s \in A^*$  is such that  $\alpha_L(u_s) = s$ . Therefore, by definition of the syntactic morphism  $\alpha_L$ , it suffices to show that  $u_s \leq_L w$ , where  $\leq_L$  is the syntactic precongruence of  $L$ . Consider  $x, y \in A^*$  such that  $xu_s y \in L$ . We have to prove that  $xwy \in L$ . Let  $q = \alpha_L(x)$  and  $r = \alpha_L(y)$ . By hypothesis, we have  $qsr = \alpha_L(xu_s y) \in \alpha_L(L)$  which yields  $(q, r) \in P_s$ . Hence, we get  $w \in u_q^{-1} L u_r^{-1}$  by hypothesis, which yields  $u_q w u_r \in L$ . Since  $\alpha_L(u_q w u_r) = q\alpha_L(w)r = \alpha_L(xwy)$  and  $L$  is recognized by  $\alpha_L$ , it then follows that  $xwy \in L$ , as desired. This completes the proof. ■

In view of Proposition 2.12, getting an algorithm for  $C$ -membership boils down to finding a procedure to decide whether an input morphism  $\alpha : A^* \rightarrow (M, \leq)$  is a  $C$ -morphism. This is how we approach the question in the paper.

Additionally, we shall use  $C$ -morphisms as mathematical tools in proof arguments. They are convenient when manipulating arbitrary classes. We present a few properties that we shall need in this context. First, we have the following simple corollary of Proposition 2.12.

**PROPOSITION 2.13.** Let  $C$  be a positive prevariety and consider finitely many languages  $L_1, \dots, L_k \in C$ . There exists a  $C$ -morphism  $\eta : A^* \rightarrow (N, \leq)$  such that  $L_1, \dots, L_k$  are recognized by  $\eta$ .

**PROOF.** For every  $i \leq k$ , we let  $\alpha_i : A^* \rightarrow (M_i, \leq_i)$  be the syntactic morphism of  $L_i$ . We know from Proposition 2.12 that  $\alpha_i$  is a  $C$ -morphism. Consider the monoid  $M = M_1 \times \dots \times M_k$  equipped with the componentwise multiplication. Moreover, consider the ordering  $\leq$  on  $M$  defined by  $(s_1, \dots, s_k) \leq (t_1, \dots, t_k)$  if and only if  $s_i \leq_i t_i$  for every  $i \leq k$ . Clearly,  $(M, \leq)$  is an ordered monoid. Moreover, let  $\alpha : A^* \rightarrow (M, \leq)$  be the morphism defined by  $\alpha(w) = (\alpha_1(w_1), \dots, \alpha_k(w))$  for every  $w \in A^*$ . One can verify from the definition that all languages recognized by  $\alpha$  are finite intersections of languages recognized by  $\alpha_1, \dots, \alpha_k$  (in particular,  $\alpha$  recognizes each  $L_i$ ). Hence, all languages recognized by  $\alpha$  belong to  $C$ . It now suffices to let  $\eta : A^* \rightarrow (N, \leq)$  be the surjective morphism obtained by restricting the codomain of  $\alpha$  to its image. ■

We now consider the particular case of  $\mathcal{G}$ -morphisms when  $\mathcal{G}$  is a prevariety of group languages. In this case, the codomain of any  $\mathcal{G}$ -morphism is a group.

**LEMMA 2.14.** Let  $\mathcal{G}$  be a prevariety of group languages and let  $\eta : A^* \rightarrow G$  be a  $\mathcal{G}$ -morphism. Then,  $G$  is a group.

**PROOF.** Let  $p = \omega(G)$ . We prove that for every  $g \in G$ , the element  $g^{p-1}$  is an inverse of  $g$ , i.e., that  $g^p = 1_G$ . Since  $\eta$  is a  $\mathcal{G}$ -morphism and  $\mathcal{G}$  is a prevariety of group languages,  $\eta^{-1}(1_G) \in \mathcal{G}$  is recognized by a finite group. Let  $\beta : A^* \rightarrow H$  be a morphism into a finite group  $H$  recognizing  $\eta^{-1}(1_G)$ . Since  $\varepsilon \in \eta^{-1}(1_G)$ , it follows that  $\beta^{-1}(1_H) \subseteq \eta^{-1}(1_G)$ . Let  $w \in \eta^{-1}(g)$  (recall that  $\eta$  is surjective, since it is a  $\mathcal{G}$ -morphism), and let  $k = \omega(H)$ . Since  $H$  is a group, we have  $(\beta(w))^{kp} = 1_H$  (the unique idempotent in  $H$ ). Therefore,  $\beta(w^{kp}) = 1_H$ , which implies that  $w^{kp} \in \eta^{-1}(1_G)$ . Hence,  $\eta(w^{kp}) = 1_G$  which exactly says that  $g^p = 1_G$  by definition of  $p$ . ■

We now look at  $\mathcal{G}^+$ -morphisms when  $\mathcal{G}$  is a prevariety of group languages.

**LEMMA 2.15.** Let  $\mathcal{G}$  be a prevariety of group languages and let  $\eta : A^* \rightarrow N$  be a  $\mathcal{G}^+$ -morphism. Then, the set  $G = \eta(A^+)$  is a group and the map  $\beta : A^* \rightarrow G$  defined by  $\beta(\varepsilon) = 1_G$  and  $\beta(w) = \eta(w)$  for  $w \in A^+$  is a  $\mathcal{G}$ -morphism.

**PROOF.** We first prove that  $G = \eta(A^+)$  is a group. Let  $p = \omega(N)$ . It suffices to show that for every  $g \in G$ , we have  $g^p = 1_G$  i.e., that for every  $g, h \in G$ , we have  $g^p h = h g^p = h$ . By definition of  $G$ , there exist  $u, v \in A^+$  such that  $\eta(u) = g$  and  $\eta(v) = h$ . By definition of  $\mathcal{G}^+$ , since  $\eta : A^* \rightarrow N$  is a  $\mathcal{G}^+$ -morphism, we know that there exists a language  $L_h \in \mathcal{G}$  such that either  $\eta^{-1}(h) = \{\varepsilon\} \cup L_h$  or  $\eta^{-1}(h) = A^+ \cap L_h$ . Since  $v \in A^+$  and  $\eta(v) = h$ , we have  $v \in L_h$ . Since  $L_h \in \mathcal{G}$  is a group language, there exists a number  $n \geq 1$  such that  $u^{np}v \in L_h$  and  $vu^{np} \in L_h$ : it suffices to choose  $n$  as the idempotent power of a finite group recognizing  $L_h$ . Since  $u^{np}v, vu^{np} \in A^+$ , it follows that  $\eta(u^{np}v) = \eta(vu^{np}) = h$ . Since  $p = \omega(N)$ , this exactly says that  $g^p h = h g^p = h$ , concluding the proof that  $G$  is a group. Finally, consider the morphism  $\beta : A^* \rightarrow G$  defined by  $\beta(\varepsilon) = 1_G$  and  $\beta(w) = \eta(w)$  for every  $w \in A^+$ . Let  $L$  be a language recognized by  $\beta$ . By definition  $L$  is a group language and it is also recognized by  $\eta$ , which implies that  $L \in \mathcal{G}^+$  by hypothesis on  $\eta$ . Hence, it follows from Lemma 2.9 that  $L \in \mathcal{G}$ , and we conclude that  $\beta$  is a  $\mathcal{G}$ -morphism. ■

Finally, we consider the special case where  $C$  is a *finite* prevariety (i.e.,  $C$  contains finitely many languages). First, recall from Lemma 2.10 that being a  $C$ -morphism does not depend on the ordering of the finite monoid, which we can therefore assume to be unordered. Moreover, since  $C$  is finite, Proposition 2.13 implies that there exists a  $C$ -morphism recognizing *all* languages in  $C$ . The following lemma implies that it is unique (up to renaming).

**LEMMA 2.16.** Let  $C$  be a finite prevariety and let  $\alpha : A^* \rightarrow M$  and  $\eta : A^* \rightarrow N$  be two  $C$ -morphisms. If  $\alpha$  recognizes all languages in  $C$ , then there exists a morphism  $\gamma : M \rightarrow N$  such that  $\eta = \gamma \circ \alpha$ .

**PROOF.** We assume that  $\alpha$  recognizes all languages in  $C$  and define  $\gamma : M \rightarrow N$ . For every  $s \in M$ , we fix a word  $w_s \in \alpha^{-1}(s)$  (recall again that  $C$ -morphisms are surjective by definition) and define  $\gamma(s) = \eta(w_s)$ . It remains to prove that  $\gamma$  is a morphism and that  $\eta = \gamma \circ \alpha$ . It suffices to prove the latter: since  $\alpha$  is surjective, the former is an immediate consequence. Let  $v \in A^*$ . We show that  $\eta(v) = \gamma(\alpha(v))$ . Let  $s = \alpha(v)$ . By definition,  $\gamma(s) = \eta(w_s)$ . Hence, we need to prove that  $\eta(v) = \eta(w_s)$ . Since  $\eta$  is a  $C$ -morphism, we have  $\eta^{-1}(\eta(w_s)) \in C$ . Hence, our hypothesis implies that  $\eta^{-1}(\eta(w_s))$  is recognized by  $\alpha$ . Since it is clear that  $w_s \in \eta^{-1}(\eta(w_s))$  and  $\alpha(v) = \alpha(w_s) = s$ , it follows that  $v \in \eta^{-1}(\eta(w_s))$  which exactly says that  $\eta(v) = \eta(w_s)$ , completing the proof. ■

By Lemma 2.16, if  $C$  is a finite prevariety and  $\alpha : A^* \rightarrow M$  and  $\eta : A^* \rightarrow N$  are two  $C$ -morphisms which both recognize *all* languages in  $C$ , there exist two morphisms  $\gamma : M \rightarrow N$  and  $\beta : N \rightarrow M$  such that  $\eta = \gamma \circ \alpha$  and  $\alpha = \beta \circ \eta$ . Since  $\alpha$  and  $\eta$  are surjective, it follows that  $\beta \circ \gamma : M \rightarrow M$  is the identity morphism. Hence,  $\beta$  and  $\gamma$  are both isomorphisms which means that  $\alpha$  and  $\eta$  are the same object up to renaming. We call it the *canonical  $C$ -morphism* and denote it by  $\eta_C : A^* \rightarrow N_C$ . Let us emphasize that this object is only defined when  $C$  is a *finite prevariety*.

**EXAMPLE 2.17.** An important example of finite prevariety is the class AT of alphabet testable languages defined above: the Boolean combinations of languages  $A^*aA^*$  for  $a \in A$ . It can be verified that  $N_{\text{AT}}$  is the monoid  $2^A$  of subalphabets whose multiplication is union (up to isomorphism). Moreover,  $\eta_{\text{AT}} : A^* \rightarrow 2^A$  is the morphism such that for every  $w \in A^*$ ,  $\eta_{\text{AT}}(w) \in 2^A$  is the set of letters occurring in  $w$  (i.e., the least  $B \subseteq A$  such that  $w \in B^*$ ).

### 3. Fragments of first-order logic

We recall the definition of first-order logic over words and its fragments. In the paper, we are interested in two particular fragments which we define here: two-variable first-order logic  $\text{FO}^2$  and level  $\Delta_2$  in the quantifier alternation hierarchy of first-order logic.

**Positions.** We view a word  $w \in A^*$  as a logical structure whose domain is a set of positions. More precisely, a word  $w = a_1 \cdots a_{|w|} \in A^*$  for the letters  $a_1, \dots, a_{|w|} \in A$ , is viewed as an *ordered set of  $|w| + 2$  positions*  $\text{Pos}(w) = \{0, 1, \dots, |w|, |w| + 1\}$ . Each position  $i$  such that  $1 \leq i \leq |w|$  carries the label  $a_i \in A$ . On the other hand, positions 0 and  $|w| + 1$  are *artificial* leftmost and rightmost positions which carry *no label*.

**REMARK 3.1.** With this convention, every word has at least two positions. In particular, the empty word  $\varepsilon$  contains only the two unlabeled positions: we have  $\text{Pos}(\varepsilon) = \{0, 1\}$ .

Given a word  $w = a_1 \cdots a_{|w|} \in A^*$  and a position  $i \in \text{Pos}(w)$ , we associate an element  $w[i] \in A \cup \{\text{min}, \text{max}\}$ . If  $i = 0$ , then  $w[i] = \text{min}$  and if  $i = |w| + 1$ , then  $w[i] = \text{max}$ . Otherwise,  $1 \leq i \leq |w|$  and we let  $w[i] = a_i$ . Additionally, if  $i, j \in \text{Pos}(w)$  are *two* positions such that  $i < j$ , we write  $w(i, j) = a_{i+1} \cdots a_{j-1} \in A^*$  (i.e., the infix obtained by keeping the letters carried by positions that are *strictly* between  $i$  and  $j$ ). Note that  $w(0, |w| + 1) = w$ .

**First-order logic.** We use first-order logic (FO) to express properties of words  $w$  in  $A^*$ . A particular formula can quantify over the positions in  $w$  and use a predetermined set of predicates to test properties of these positions. We also allow two constants “*min*” and “*max*” interpreted as the artificial unlabeled positions 0 and  $|w| + 1$ . For a formula  $\varphi(x_1, \dots, x_n)$  with free variables  $x_1, \dots, x_n$ , a word  $w \in A^*$  and positions  $i_1, \dots, i_n \in \text{Pos}(w)$ , we write  $w \models \varphi(i_1, \dots, i_n)$  to indicate that  $w$  satisfies  $\varphi$  when  $x_1, \dots, x_n$  are interpreted as the positions  $i_1, \dots, i_n$ . A sentence  $\varphi$  is a formula without free variables. It defines the language  $L(\varphi) = \{w \in A^* \mid w \models \varphi\}$ .

We use two kinds of predicates. The first kind is standard. For each  $a \in A$ , we have a unary predicate (also denoted by  $a$ ) selecting all positions labeled by “ $a$ ”. We also use three binary predicates: equality “ $=$ ”, the (strict) linear order “ $<$ ” and the successor “ $+1$ ”.

**EXAMPLE 3.2.** The language  $A^*abA^*cA^*d$  is defined by the following FO sentence:

$$\exists x_1 \exists x_2 \exists x_3 \exists x_4 (x_1 + 1 = x_2) \wedge (x_2 < x_3) \wedge (x_4 + 1 = \text{max}) \wedge a(x_1) \wedge b(x_2) \wedge c(x_3) \wedge d(x_4).$$

A *fragment* of first-order logic consists in the specification of a (possibly finite) set  $V$  of variables and a set  $\mathcal{F}$  of first-order formulas using only the variables of  $V$ . This set  $\mathcal{F}$  must contain all quantifier-free formulas and must be closed under disjunction, conjunction and quantifier-free substitution (if  $\varphi \in \mathcal{F}$ , replacing a quantifier-free subformula in  $\varphi$  by another quantifier-free formula produces a new formula in  $\mathcal{F}$ ). If  $\mathcal{S}$  is a set of predicates and  $\mathcal{F}$  is a fragment of first-order logic, we write  $\mathcal{F}(\mathcal{S})$  for the class containing all languages  $L(\varphi)$  where  $\varphi$  is a sentence of  $\mathcal{F}$  using only predicates that belong to  $\mathcal{S} \cup A \cup \{=\}$  (the label predicates and the equality predicate are *always* implicitly allowed). We provide examples at the end of the section (“*Particular fragments*”).

We consider generic *signatures*, *i.e.*, possibly infinite sets of predicates built from an arbitrary input class  $C$ . There are of two kinds:

- First, we consider the set  $\mathbb{I}_C$  containing a binary “infix” predicate  $I_L(x, y)$  for every language  $L \in C$ . Given a word  $w \in A^*$  and two positions  $i, j \in \text{Pos}(w)$ , we have  $w \models I_L(i, j)$  if and only if  $i < j$  and  $w(i, j) \in L$ .
- Second, we consider the set  $\mathbb{P}_C$  containing a unary “prefix” predicate  $P_L(x)$  for every language  $L \in C$ . Given a word  $w \in A^*$  and a position  $i \in \text{Pos}(w)$ , we have  $w \models P_L(i)$  if and only if  $0 < i$  and  $w(0, i) \in L$ .

Observe that the predicates available in  $\mathbb{P}_C$  are easily expressed from those in  $\mathbb{I}_C$ : clearly, the formula  $P_L(x)$  is equivalent to  $I_L(\text{min}, x)$ .

The sets of predicates  $\mathbb{P}_C$  and  $\mathbb{I}_C$  are defined for every class  $C$ . Yet, we shall mostly be interested in the cases when  $C$  is either a prevariety of group languages  $\mathcal{G}$  or its well-suited extension  $\mathcal{G}^+$ . This is motivated by the following lemma.

**LEMMA 3.3.** Let  $\mathcal{G}$  be a prevariety of group languages and let  $\mathcal{F}$  be a fragment of first-order logic. Then,  $\mathcal{F}(\mathbb{I}_{\mathcal{G}}) = \mathcal{F}(<, \mathbb{P}_{\mathcal{G}})$  and  $\mathcal{F}(\mathbb{I}_{\mathcal{G}^+}) = \mathcal{F}(<, +1, \mathbb{P}_{\mathcal{G}})$ .

**PROOF.** We first prove that  $\mathcal{F}(<, \mathbb{P}_{\mathcal{G}}) \subseteq \mathcal{F}(\mathbb{I}_{\mathcal{G}})$  and  $\mathcal{F}(<, +1, \mathbb{P}_{\mathcal{G}}) \subseteq \mathcal{F}(\mathbb{I}_{\mathcal{G}^+})$ . It suffices to prove that we may express all atomic formulas of  $\mathcal{F}(<, \mathbb{P}_{\mathcal{G}})$  and  $\mathcal{F}(<, +1, \mathbb{P}_{\mathcal{G}})$  using atomic formulas of  $\mathcal{F}(\mathbb{I}_{\mathcal{G}})$  and  $\mathcal{F}(\mathbb{I}_{\mathcal{G}^+})$  respectively. The linear order  $x < y$  is expressed by  $I_{A^*}(x, y)$ . For every  $L \in \mathcal{G}$ ,  $P_L(x)$  is expressed by  $I_L(\text{min}, x)$ . Finally, the successor relation  $x + 1 = y$  is expressed by  $I_{\{\varepsilon\}}(x, y)$  (by definition of  $\mathcal{G}^+$ , we know that  $I_{\{\varepsilon\}}$  is a predicate of  $\mathbb{I}_{\mathcal{G}^+}$ ).

We now prove that  $\mathcal{F}(\mathbb{I}_{\mathcal{G}}) \subseteq \mathcal{F}(<, \mathbb{P}_{\mathcal{G}})$ . By definition of fragments, it suffices to prove that for every  $L \in \mathcal{G}$ , the atomic formula  $I_L(x, y)$  is equivalent to a quantifier-free formula of  $\mathcal{F}(<, \mathbb{P}_{\mathcal{G}})$ . Proposition 2.13 yields a  $\mathcal{G}$ -morphism  $\alpha : A^* \rightarrow G$  recognizing  $L$ . Since  $\mathcal{G}$  is a prevariety of group languages,  $G$  is a group by Lemma 2.14. Since  $\alpha$  is a  $\mathcal{G}$ -morphism, the language  $\alpha^{-1}(g)$  belongs to  $\mathcal{G}$  for every  $g \in G$ , whence  $P_{\alpha^{-1}(g)}$  is a predicate available in  $\mathbb{P}_{\mathcal{G}}$ . Let  $F \subseteq G$  be the set such that  $\alpha^{-1}(F) = L$ . Since  $G$  is a group, we have  $\alpha(v) = (\alpha(ua))^{-1}\alpha(uav)$  for all  $u, v \in A^*$  and  $a \in A$ . We define  $T = \{(g, a, h) \in G \times A \times G \mid (g\alpha(a))^{-1}h \in F\}$ . Consider the

following quantifier-free formula of  $\mathcal{F}(<, \mathbb{P}_{\mathcal{G}})$ :

$$\varphi(x, y) = (x < y) \wedge \left( \bigvee_{(g,a,h) \in T} (P_{\alpha^{-1}(g)}(x) \wedge a(x) \wedge P_{\alpha^{-1}(h)}(y)) \right).$$

One now may verify that the atomic formula  $I_L(x, y)$  is equivalent to the following quantifier-free formula of  $\mathcal{F}(<, \mathbb{P}_{\mathcal{G}})$ :

$$(x = \min \wedge P_L(y)) \vee \varphi(x, y).$$

This concludes the proof of the inclusion  $\mathcal{F}(\mathbb{I}_{\mathcal{G}}) \subseteq \mathcal{F}(<, \mathbb{P}_{\mathcal{G}})$ .

Finally, we prove that  $\mathcal{F}(\mathbb{I}_{\mathcal{G}^+}) \subseteq \mathcal{F}(<, +1, \mathbb{P}_{\mathcal{G}})$ . By definition, it suffices to show that for every language  $K \in \mathcal{G}^+$ , the atomic formula  $I_K(x, y)$  is equivalent to a quantifier-free formula of  $\mathcal{F}(<, +1, \mathbb{P}_{\mathcal{G}})$ . By definition of  $\mathcal{G}^+$ , there exists  $L \in \mathcal{G}$  such that either  $K = \{\varepsilon\} \cup L$  or  $K = A^+ \cap L$ . Consequently,  $I_K(x, y)$  is equivalent to either  $I_{\{\varepsilon\}}(x, y) \vee I_L(x, y)$  or  $I_{A^+}(x, y) \wedge I_L(x, y)$ . Since,  $L \in \mathcal{G}$ , we already proved above that  $I_L(x, y)$  is equivalent to a quantifier-free formula of  $\mathcal{F}(<, \mathbb{P}_{\mathcal{G}}) \subseteq \mathcal{F}(<, +1, \mathbb{P}_{\mathcal{G}})$ . Moreover,  $I_{\{\varepsilon\}}(x, y)$  is equivalent to  $x + 1 = y$  and  $I_{A^+}$  is equivalent to  $x < y \wedge \neg(x + 1 = y)$ . This concludes the proof. ■

Lemma 3.3 covers many natural sets of predicates. We present three important cases. If  $\mathcal{G}$  is the trivial prevariety  $\text{ST} = \{\emptyset, A^*\}$ , all predicates in  $\mathbb{P}_{\text{ST}}$  are trivial. Hence, we get the classes  $\mathcal{F}(<)$  and  $\mathcal{F}(<, +1)$ . For the class MOD of *modulo languages* (see Section 2.3 for its definition), one can check that in this case, we obtain the classes  $\mathcal{F}(<, \text{MOD})$  and  $\mathcal{F}(<, +1, \text{MOD})$  where “MOD” is the set of *modular predicates*, defined as follows. For all  $k, m \in \mathbb{N}$  such that  $k < m$ , the set MOD contains a unary predicate  $M_{k,m}$  selecting all positions  $i$  such that  $i \equiv k \pmod{m}$ . Finally, we may consider the class AMT of *alphabet modulo testable languages* (see Section 2.3 for its definition). In this case, we get the classes  $\mathcal{F}(<, \text{AMOD})$  and  $\mathcal{F}(<, +1, \text{AMOD})$  where “AMOD” is the set of *alphabetic modular predicates*: for all  $a \in A$  and  $k, m \in \mathbb{N}$  such that  $k < m$ , the set AMOD contains a unary predicate  $M_{k,m}^a$  selecting all positions  $i$  such  $|w(0, i)|_a \equiv k \pmod{m}$ .

**Particular fragments.** In the paper, we consider *two* special fragments. The first one is *two-variable first-order logic* ( $\text{FO}^2$ ). It consists of all first-order formulas that use at most *two* distinct variables (which can be reused). In the formal definition, introducing the fragment  $\text{FO}^2$  boils down to using a set  $V$  of variables which has size two. For example, the sentence

$$\exists x \exists y x < y \wedge M_{1,2}(x) \wedge a(x) \wedge b(y) \wedge (\exists x y < x \wedge c(x))$$

is an  $\text{FO}^2$  sentence using the label predicates, the linear order predicate and the modular predicate “ $M_{1,2}$ ”. It defines the language  $(AA)^* aA^* bA^* cA^*$ , which therefore belongs to the class  $\text{FO}^2(<, \text{MOD})$ .

We also consider individual levels in the quantifier alternation hierarchy of full first-order logic. One may classify first-order sentences by counting the alternations between  $\exists$  and  $\forall$  quantifiers in their prenex normal form. For  $n \geq 1$ , an FO sentence is  $\Sigma_n$  (resp.  $\Pi_n$ ) when its

prenex normal form has  $(n - 1)$  quantifier alternations (that is,  $n$  blocks of quantifiers) and starts with an  $\exists$  (resp. a  $\forall$ ) quantifier. For example, a sentence whose prenex normal form is,

$$\exists x_1 \exists x_2 \forall x_3 \exists x_4 \varphi(x_1, x_2, x_3, x_4) \quad (\text{with } \varphi \text{ quantifier-free})$$

is  $\Sigma_3$ . Observe that the sets of  $\Sigma_n$  and  $\Pi_n$  sentences are not closed under negation: negating a  $\Sigma_n$  sentence yields a  $\Pi_n$  sentence and vice versa. Thus, one also considers  $\mathcal{B}\Sigma_n$  sentences: Boolean combinations of  $\Sigma_n$  sentences.

We are interested in intermediary levels that are not defined directly by a set of formulas. If  $\mathcal{S}$  is a set of predicates and  $n \geq 1$ , we may consider the classes  $\Sigma_n(\mathcal{S})$  and  $\Pi_n(\mathcal{S})$ . It is also standard to consider a third class, denoted by  $\Delta_n(\mathcal{S})$ : it consists exactly of the languages that can be defined by both a  $\Sigma_n$  formula *and* a  $\Pi_n$  formula (using the predicates in  $\mathcal{S} \cup A \cup \{=\}$ ). In other words,  $\Delta_n(\mathcal{S}) = \Sigma_n(\mathcal{S}) \cap \Pi_n(\mathcal{S})$ . In the paper, we look at the classes  $\Delta_2(\mathbb{I}_C)$ .

## 4. Operators

In this section, we introduce the operators investigated in the paper. They are all variants of the *polynomial closure operator*  $Pol$ , which we first define. Let us point out that the paper is not about the operator  $Pol$  itself: all results about  $Pol$  that we use are taken from previous work. Then, we introduce *unambiguous polynomial closure* ( $UPol$ ), which is a semantic restriction of  $Pol$ . This is the main operator of the paper. Finally, we present another variant of  $Pol$ : *alternated polynomial closure* ( $APol$ ). We use it as a tool for investigating  $UPol$ . In particular, we shall prove that  $UPol(C) = APol(C)$  when  $C$  is a prevariety.

### 4.1 Polynomial closure

The definition is based on *marked* concatenation. Abusing terminology, given a word  $u \in A^*$ , we denote by  $u$  the singleton language  $\{u\}$ . Given two languages  $K, L \subseteq A^*$ , a *marked concatenation* of  $K$  with  $L$  is a language of the form  $KaL$ , for some letter  $a \in A$ .

**DEFINITION 4.1.** Consider a class  $C$ . The *polynomial closure* of  $C$ , denoted by  $Pol(C)$  is the least class containing  $C$  closed under union and marked concatenation. That is, for every  $K, L \in Pol(C)$  and  $a \in A$ , we have  $K \cup L \in Pol(C)$  and  $KaL \in Pol(C)$ .

**EXAMPLE 4.2.** Consider the class  $ST = \{\emptyset, A^*\}$ . Then, using the fact that language concatenation distributes over union, it is easy to check that  $Pol(ST)$  consists of all finite unions of languages  $A^*a_1A^* \cdots a_nA^*$  with  $a_1, \dots, a_n \in A$ .

It is not clear from the definition whether the class  $Pol(C)$  has robust properties, even when  $C$  does. It was shown by Arfi [4, 3] that if  $C$  is a prevariety, then  $Pol(C)$  is a positive prevariety (the key issue is closure under intersection, which is not required by the definition).

This result was later strengthened by Pin [16]: if  $C$  is a positive prevariety, then so is  $Pol(C)$  (see also [26] for a more recent proof).

**THEOREM 4.3** (Arfi [4, 3], Pin [16]). *Let  $C$  be a positive prevariety. Then,  $Pol(C)$  is a positive prevariety as well.*

In general, the classes  $Pol(C)$  built with polynomial closure are *not* closed under complement, even when the class  $C$  is (see  $Pol(ST)$  in Example 4.2). Consequently, it is natural to combine polynomial closure with two other independent operators. The first one is *Boolean closure*. Given an input class  $\mathcal{D}$ , we write  $Bool(\mathcal{D})$  for the *smallest* Boolean algebra containing  $\mathcal{D}$ . Moreover, given an input class  $C$ , we write  $BPol(C)$  for  $Bool(Pol(C))$ .

**EXAMPLE 4.4.** In view of Example 4.2,  $BPol(ST)$  consists of all Boolean combinations of languages  $A^*a_1A^*\cdots a_nA^*$  with  $a_1, \dots, a_n \in A$ . This is the class of piecewise testable languages [36], which is prominent in the literature.

**EXAMPLE 4.5.** It can be verified that  $BPol(AT)$  consists exactly of all Boolean combinations of marked products  $B_0^*a_1B_1^*\cdots a_nB_n^*$  with  $a_1, \dots, a_n \in A$  and  $B_0, \dots, B_n \subseteq A$ . It is known [21], that this is exactly the class  $BPol(BPol(ST))$ , as we shall see in Example 4.19 below.

Since one quotients commute with Boolean operations, the following statement is an immediate corollary of Theorem 4.3.

**COROLLARY 4.6.** Let  $C$  be a positive prevariety. Then,  $BPol(C)$  is a prevariety.

The second operator is defined as follows. Given a class  $C$ , we write  $coPol(C)$  for the class containing all complements of languages in  $Pol(C)$ . That is,

$$coPol(C) = \{A^* \setminus L \mid L \in Pol(C)\}.$$

We shall consider the class  $Pol(C) \cap coPol(C)$  consisting of all languages that belong to both  $Pol(C)$  and  $coPol(C)$ . Consequently,

$$Pol(C) \cap coPol(C) = \{L \mid L \in Pol(C) \text{ and } A^* \setminus L \in Pol(C)\}.$$

In other words, since  $Pol(C)$  is closed under union and intersection,  $Pol(C) \cap coPol(C)$  is the *greatest* Boolean algebra contained in  $Pol(C)$ . We have the following immediate corollary of Theorem 4.3.

**COROLLARY 4.7.** Let  $C$  be a positive prevariety. Then,  $coPol(C)$  is a positive prevariety and  $Pol(C) \cap coPol(C)$  is a prevariety.

Finally, we have the following useful result about classes of the form  $Pol(BPol(C))$  which is taken from [26]. We recall the proof, which is straightforward.

**LEMMA 4.8.** Let  $C$  be a prevariety. Then,  $Pol(BPol(C)) = Pol(coPol(C))$ .

**PROOF.** It is clear that  $coPol(C) \subseteq BPol(C)$ , whence  $Pol(coPol(C)) \subseteq Pol(BPol(C))$ . We show that  $BPol(C) \subseteq Pol(coPol(C))$ . It will follow that,

$$Pol(BPol(C)) \subseteq Pol(Pol(coPol(C))) = Pol(coPol(C)).$$

Let  $L \in BPol(C)$ . By definition,  $L$  is a Boolean combination of languages in  $Pol(C)$ . Hence, De Morgan's laws show that  $L$  is built by applying unions and intersections to languages that are elements of either  $Pol(C)$  or  $coPol(C)$ . Clearly,  $Pol(coPol(C))$  contains both  $Pol(C)$  and  $coPol(C)$ . Moreover, it follows from Theorem 4.3 and Corollary 4.7 that  $Pol(coPol(C))$  is a positive prevariety. Altogether, we obtain  $L \in Pol(coPol(C))$  as desired. ■

**Logical characterizations.** It is well known that the quantifier alternation hierarchies of first-order logic can be characterized using the operators  $Pol$  and  $Bool$ . Historically, this was first proved by Thomas [41].

Given a single input class  $C$ , the *concatenation hierarchy of basis  $C$*  is built from by applying  $Pol$  and  $Bool$  to  $C$  in alternation. More precisely, the hierarchy is made of two kinds of levels. The *full levels* are denoted by natural numbers  $0, 1, 2, 3, \dots$ . The *half levels* are denoted by  $1/2, 3/2, 5/2, \dots$ . Level 0 is the basis  $C$ . Then, for every number  $n \in \mathbb{N}$ :

- the half level  $n + 1/2$  is the polynomial closure ( $Pol$ ) of level  $n$ .
- the full level  $n + 1$  is the Boolean closure ( $Bool$ ) of level  $n + 1/2$ .

When  $C$  is a prevariety, this hierarchy exactly characterizes the quantifier alternation hierarchy of first-order logic associated to the set of predicates  $\mathbb{I}_C$  (see [26]). More precisely, for all  $n \geq 1$ , level  $n - 1/2$  is exactly the class  $\Sigma_n(\mathbb{I}_C)$  and level  $n$  is exactly the class  $\mathcal{B}\Sigma_n(\mathbb{I}_C)$ .

We are mainly interested in the intermediary levels  $\Delta_n(\mathbb{I}_C) = \Sigma_n(\mathbb{I}_C) \cap \Pi_n(\mathbb{I}_C)$  and more precisely in the special case where  $n = 2$ . As explained above,  $\Sigma_2(\mathbb{I}_C) = Pol(BPol(C))$  and one can check that this implies  $\Pi_2(\mathbb{I}_C) = coPol(BPol(C))$ . We get the following theorem.

**THEOREM 4.9.** For every prevariety  $C$ , we have  $\Delta_2(\mathbb{I}_C) = Pol(BPol(C)) \cap coPol(BPol(C))$ .

## 4.2 Unambiguous polynomial closure

We turn to the main operator of the paper. It is defined as a variant of polynomial closure obtained by restricting marked concatenations to *unambiguous* ones and unions to *disjoint ones*. We first define what it means for some marked concatenation to be unambiguous. In fact, we introduce three particular kinds of marked concatenations.

**Deterministic marked concatenations.** We define three properties. Let  $K, L \subseteq A^*$  and  $a \in A$ . Consider the marked concatenation  $KaL$ . We say that,

- $KaL$  is *left deterministic* when  $K \cap KaA^* = \emptyset$ .

- $KaL$  is *right deterministic* when  $L \cap A^*aL = \emptyset$ .
- $KaL$  is *unambiguous* when for every  $w \in KaL$ , there exists a *unique* decomposition of  $w$  witnessing this membership. In other words,  $KaL$  is unambiguous if for every  $u, u' \in K$  and  $v, v' \in L$  such that  $uav = u'av'$ , we have  $u = u'$  and  $v = v'$ .

**EXAMPLE 4.10.** Let  $A = \{a, b, c\}$ . Then  $b^*aA^*$  is left deterministic and  $A^*ab^*$  is right deterministic. Moreover, they are both unambiguous. On the other hand,  $(ab)^+a(ca)^+$  is unambiguous but is neither left, nor right deterministic. Finally,  $A^*aA^*$  is ambiguous:  $aa \in A^*aA^*$  and there are two decompositions of  $aa$  witnessing this membership.

**REMARK 4.11.** Marked concatenations of nonempty group languages are ambiguous. Indeed, let  $K, L \subseteq A^*$  be two nonempty group languages and let  $a \in A$ . Let  $\alpha : A^* \rightarrow G$  be a morphism into a finite group  $G$  recognizing both  $K$  and  $L$ , and let  $p = \omega(G) > 0$ . Moreover, let  $u \in K$  and  $v \in L$ . Then  $\alpha(ua^p) = \alpha(u)$ , whence  $ua^p \in K$ . Similarly,  $a^pv \in L$ . Therefore, the word  $(ua^p)av = ua(a^pv)$  has two decompositions witnessing its membership in  $KaL$ .

**REMARK 4.12.** Being deterministic or unambiguous is a semantic property: whether  $KaL$  satisfies it may not be apparent on the definitions of  $K$  and  $L$ . Moreover, this depends on  $K, L$  and  $a$ , not just on the concatenated language  $KaL$ . There may exist two marked concatenations representing the same language and such that one is unambiguous while the other is not. For example, if  $A = \{a, b\}$ , it is clear that  $b^*aA^* = A^*aA^*$ . However,  $b^*aA^*$  is left deterministic and unambiguous while  $A^*aA^*$  is neither.

**REMARK 4.13.** It is immediate from the definitions that if the marked concatenation  $KaL$  is unambiguous (resp. left deterministic, right deterministic), then for every  $K' \subseteq K$  and  $L' \subseteq L$ , the marked concatenation  $K'aL'$  is also unambiguous (resp. left deterministic, right deterministic). We shall often use this simple fact implicitly.

An alternative way to introduce left/right deterministic marked concatenations is to present them as special kinds of unambiguous marked concatenations. We detail this point in the following lemma.

**LEMMA 4.14.** Let  $K, L \subseteq A^*$  and  $a \in A$ . The two following properties hold:

- $KaL$  is left deterministic if and only if  $KaA^*$  is unambiguous.
- $KaL$  is right deterministic if and only if  $A^*aL$  is unambiguous

**PROOF.** By symmetry, we only prove the first assertion. We prove its contrapositive:  $KaL$  is *not* left deterministic if and only if  $KaA^*$  is *not* unambiguous. Assume first that  $KaL$  is not left deterministic. This yields  $u \in K \cap KaA^*$ . Hence, we get  $u' \in K$  and  $v \in A^*$  such that  $u = u'av$ . Hence,  $ua = u'ava \in KaA^*$  and there are two decompositions witnessing this fact:  $(u)a(\varepsilon)$  and

$(u')a(va)$ . We conclude that  $KaA^*$  is ambiguous. Conversely, assume that  $KaA^*$  is ambiguous. We get  $u, u' \in K$  and  $v, v' \in A^*$  such that  $u \neq u'$  and  $uav = u'av'$ . By hypothesis, either  $ua$  is a prefix of  $u'$  or  $u'a$  is a prefix of  $u$ . By symmetry, we consider the case where  $ua$  is a prefix of  $u'$ . This yields  $x \in A^*$  such that  $u' = uax$ . Hence,  $u' \in K \cap KaA^*$  and we conclude that  $KaL$  is not left deterministic, as desired. ■

Note that Lemma 4.14 and Remark 4.13 imply that left (resp. right) deterministic marked concatenations are unambiguous. The converse is not true, as shown by Example 4.10. We complete the definition with a lemma connecting left (resp. right) deterministic marked concatenations to the Green relation  $\mathcal{R}$  (resp.  $\mathcal{L}$ ) defined on monoids. We shall use it in order to prove that marked concatenations are left or right deterministic.

**LEMMA 4.15.** Let  $\alpha : A^* \rightarrow N$  be a morphism into a finite monoid. Let  $s \in N$ ,  $a \in A$  and  $L \subseteq \alpha^{-1}(s)$ . The two following properties hold:

- if  $s\alpha(a) <_{\mathcal{R}} s$ , then  $LaA^*$  is left deterministic.
- if  $\alpha(a)s <_{\mathcal{L}} s$ , then  $A^*aL$  is right deterministic.

**PROOF.** By symmetry, we only prove the first assertion. We assume that  $s\alpha(a) <_{\mathcal{R}} s$  and show that  $LaA^*$  is left deterministic. By contradiction, assume that there exists a word  $w \in L \cap LaA^*$ . Since  $w \in LaA^*$ , we get  $u \in L$  and  $v \in A^*$  such that  $w = uav$ . Since  $u, w \in L$  and  $L \subseteq \alpha^{-1}(s)$ , we have  $\alpha(u) = \alpha(w) = s$ . Therefore,  $s = s\alpha(a)\alpha(v)$  which yields  $s \leq_{\mathcal{R}} s\alpha(a)$ , contradicting the hypothesis that  $s\alpha(a) <_{\mathcal{R}} s$ . ■

**DEFINITION 4.16 (Unambiguous polynomial closure).** Consider a class  $C$ . Its *unambiguous polynomial closure*, denoted by  $UPol(C)$  is the least class containing  $C$  and closed under both *disjoint union* and *unambiguous marked concatenation*. That is given  $K, L \in UPol(C)$  and  $a \in A$ , if  $K \cap L = \emptyset$ , then  $K \uplus L \in UPol(C)$  and if  $KaL$  is unambiguous, then  $KaL \in UPol(C)$ . Note that here, we choose to use the symbol “ $\uplus$ ” for union in order to emphasize disjointedness.

**REMARK 4.17.** It is clear from the definition that  $UPol(C) \subseteq Pol(C)$  for every class  $C$ . In general, this inclusion is strict. This will be the case for all examples considered in the paper. For example, consider the class  $ST = \{\emptyset, A^*\}$ . We have  $UPol(ST) = ST$  (this is because marked concatenations  $A^*aA^*$  are ambiguous). Hence,  $UPol(ST) \subsetneq Pol(ST)$ .

**EXAMPLE 4.18.** As seen in Remark 4.17, we have  $UPol(ST) = ST$ . This is actually a generic property of prevarieties of group languages: for every such class  $\mathcal{G}$ , we have  $UPol(\mathcal{G}) = \mathcal{G}$ . This is because the marked concatenation of two nonempty group languages can never be unambiguous, see Remark 4.11. This illustrates a particularity of  $UPol$ : it is meant to be applied to “complex enough” classes.

**EXAMPLE 4.19.** An important example is  $UPol(AT)$ . It is simple to verify that this class contains exactly the finite disjoint unions of unambiguous marked products  $B_0^*a_1B_1^* \cdots a_nB_n^*$  where  $a_1, \dots, a_n \in A$  and  $B_0, \dots, B_n \subseteq A$  (unambiguous marked products are defined in the natural way by generalizing the definition to products of arbitrarily many languages). It is known as the class of *unambiguous languages* [35]. We prove below that  $UPol(AT) = UPol(BPol(ST))$ . Note that it follows from this equality that  $Pol(AT) = Pol(BPol(ST))$  and that  $BPol(AT) = BPol(BPol(ST))$ , as stated in Example 4.5.

**REMARK 4.20.** In the literature, a less restrictive definition, which allows arbitrary finite unions, is sometimes used (see [16] for example). It turns out that this is equivalent when the input class  $C$  is a prevariety. This is the only case that we consider in the paper. Yet, allowing arbitrary finite unions is important when the input class  $C$  is not closed under complement.

We now present properties of unambiguous polynomial closure that we shall prove in the paper. We start with an important theorem, which generalizes a result by Pin and Weil [20]. It turns out that when the input class  $C$  is a prevariety<sup>3</sup>, the classes  $UPol(C)$  and  $Pol(C) \cap coPol(C)$  coincide.

**THEOREM 4.21.** *For every prevariety  $C$ , we have  $UPol(C) = Pol(C) \cap coPol(C)$ .*

We postpone the proof of Theorem 4.21 to Section 7. It is based on algebraic characterizations: we prove independently that for every prevariety  $C$ , the classes  $UPol(C)$  and  $Pol(C) \cap coPol(C)$  have the same algebraic characterization. Hence, they are equal.

**REMARK 4.22.** Theorem 4.21 fails if  $C$  is only a positive prevariety (i.e., when  $C$  is not closed under complement). For example, let  $C = Pol(ST)$ . Clearly, we have  $UPol(C) = C$ . On the other hand, one can verify that  $Pol(C) \cap coPol(C) = Pol(ST) \cap coPol(ST) = ST$ . Hence, we have the strict inclusion  $Pol(C) \cap coPol(C) \subsetneq UPol(C)$  in this case.

**REMARK 4.23.** In view of Theorem 4.9, a consequence of Theorem 4.21 is that for every prevariety  $C$ , we have  $UPol(BPol(C)) = \Delta_2(\mathbb{I}_C)$ . Hence, we get a first logical characterization of  $UPol$ . We come back to this point in Section 10 where we present a second characterization of  $UPol$  in terms of two-variable first-order logic (note however that in this case, we need stronger hypotheses on the input class  $C$ ).

With Theorem 4.21 in hand, it is simple to describe the closure properties of  $UPol(C)$ . By Corollary 4.7, we know that when  $C$  is a prevariety, then so is  $Pol(C) \cap coPol(C)$ . In view of Theorem 4.21, we obtain the same result for  $UPol(C)$ . This is surprising since closure under

---

<sup>3</sup> The original result of Pin and Weil only applies to input classes satisfying stronger properties involving closure under inverse images of morphisms.

Boolean operations is not apparent on the definition of  $UPol(C)$ . This generalizes a result<sup>4</sup> by Pin, Straubing and Thérien [19].

**THEOREM 4.24.** *For every prevariety  $C$ , the class  $UPol(C)$  is a prevariety as well.*

We may use Theorem 4.24 to prove a property for a particular input class  $C$  that will be important for the applications of our results. As mentioned in Example 4.19, we have the equality  $UPol(BPol(ST)) = UPol(AT)$ .

**LEMMA 4.25.** *We have  $UPol(BPol(ST)) = UPol(AT)$ .*

**PROOF.** It is immediate from the definitions that  $AT \subseteq BPol(ST)$ . Consequently, we have  $UPol(AT) \subseteq UPol(BPol(ST))$ . For the converse inclusion, we show that  $Pol(ST) \subseteq UPol(AT)$ . Since  $UPol(AT)$  is a Boolean algebra by Theorem 4.24, this yields  $BPol(ST) \subseteq UPol(AT)$  and therefore  $UPol(BPol(ST)) \subseteq UPol(UPol(AT)) = UPol(AT)$ . We now concentrate on proving that  $Pol(ST) \subseteq UPol(AT)$ . One may verify from the definition that  $Pol(ST)$  contains exactly the finite unions of languages  $A^*a_1 \cdots A^*a_nA^*$  with  $a_1, \dots, a_n \in A$ . Hence, since  $UPol(AT)$  is a prevariety by Theorem 4.24, it suffices to prove that  $A^*a_1 \cdots A^*a_nA^* \in UPol(AT)$  for every  $a_1, \dots, a_n \in A$ . For  $0 \leq i < n$ , we let  $B_i = A \setminus \{a_{i+1}\}$ . Clearly,  $B_i^* \in AT$  for every  $i < n$ . Moreover, one can check that,

$$A^*a_1 \cdots A^*a_nA^* = B_0^*a_1 \cdots B_{n-1}^*a_nA^*.$$

It can then be verified that the marked concatenations  $B_i^*a_{i+1}(B_{i+1}^*a_{i+2} \cdots B_{n-1}^*a_nA^*)$  are all unambiguous for  $0 \leq i < n$  (they are actually left deterministic). Hence, an immediate induction yields  $B_0^*a_1 \cdots B_{n-1}^*a_nA^* \in UPol(AT)$  and we get  $A^*a_1 \cdots A^*a_nA^* \in UPol(AT)$ , completing the proof. ■

We conclude the presentation with a characteristic property of the classes  $UPol(C)$  when  $C$  is an arbitrary prevariety. We use it in Section 7 to prove the algebraic characterization of unambiguous polynomial closure.

**PROPOSITION 4.26.** *Let  $C$  be a prevariety and  $L \in UPol(C)$ . There exists a  $C$ -morphism  $\eta : A^* \rightarrow N$  and  $k \in \mathbb{N}$  satisfying the following property:*

$$\text{for all } u, v, v', x, y \in A^*, \text{ if } \eta(u) = \eta(v) = \eta(v'), \text{ then } xu^kvvu^ky \in L \Leftrightarrow xu^kv'u^ky \in L. \quad (1)$$

**PROOF.** By definition, there exists a *finite* set of languages  $\mathbf{H} \subseteq C$  such that  $L$  is built from the languages in  $\mathbf{H}$  using only unambiguous marked concatenations and disjoint unions. Proposition 2.13 yields a  $C$ -morphism  $\eta : A^* \rightarrow N$  recognizing every  $H \in \mathbf{H}$ . We now use induction on the construction of  $L$  from the languages in  $\mathbf{H}$  to prove that there exists  $k \in \mathbb{N}$  such that (1)

---

4 As for Theorem 4.21, the original result of Pin, Straubing and Thérien only applies to input classes satisfying stronger properties involving closure under inverse images by morphisms.

holds. Assume first that  $L \in \mathbf{H}$  (which means that  $L$  is recognized by  $\eta$ ). We show that (1) holds for  $k = 0$ . Indeed, let  $u, v, v', x, y \in A^*$  such that  $\eta(u) = \eta(v) = \eta(v')$ . It follows that  $\eta(xu^0vu^0y) = \eta(xu^0v'u^0y)$ . Since  $L$  is recognized by  $\eta$ , this yields  $xu^0vu^0y \in L \Leftrightarrow xu^0v'u^0y \in L$  as desired.

We now assume that  $L \notin \mathbf{H}$ . In this case, there exist  $L_1, L_2$ , built from the languages in  $\mathbf{H}$  using unambiguous marked concatenations and disjoint unions such that either  $L = L_1 \uplus L_2$ , or  $L = L_1aL_2$  for some  $a \in A$  and  $L_1aL_2$  is unambiguous. Using induction, for  $i = 1, 2$ , we get  $k_i \in \mathbb{N}$  such that  $L_i$  satisfies (1) for  $k = k_i$ . If  $L = L_1 \uplus L_2$ , it is immediate that  $L$  satisfies (1) for  $k = \max(k_1, k_2)$ . We focus on the case  $L = L_1aL_2$ . Since  $L_1$  and  $L_2$  are regular (as they belong to  $UPol(C)$ ), there exists a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid  $M$  recognizing them both. Let  $p = \omega(M) \geq 1$  be the idempotent power of this monoid. Moreover, let  $h = \max(k_1, k_2)$  and  $k = p + h$ . We prove that  $L = L_1aL_2$  satisfies (1) for this number  $k$ . Let  $u, v, v', x, y \in A^*$  such that  $\eta(u) = \eta(v) = \eta(v')$ . Assume that  $xu^kvu^ky \in L$ , we prove that  $xu^kv'u^ky \in L$  (the converse implication is symmetrical).

Since  $L = L_1aL_2$ , we have  $w_1 \in L_1$  and  $w_2 \in L_2$  such that  $xu^kvu^ky = w_1aw_2$ . Using the hypothesis that  $L_1aL_2$  is unambiguous, we prove the following fact.

**LEMMA 4.27.** Either  $xu^kvu^h$  is a prefix of  $w_1$  or  $u^hvu^ky$  is a suffix of  $w_2$ .

**PROOF.** By contradiction, we assume that  $xu^kvu^h$  is not a prefix of  $w_1$  and  $u^hvu^ky$  is not suffix of  $w_2$ . Since  $k = p + h$ , we have  $w_1aw_2 = xu^kvu^ky = xu^pu^hvu^hu^py$ . Therefore, our hypothesis yields that  $u^py$  is a suffix of  $w_2$  and  $xu^p$  is a prefix of  $w_1$ . We get  $z_1, z_2 \in A^*$  such that  $w_1 = xu^pz_1$  and  $w_2 = z_2u^py$ . Since  $w_1aw_2 = xu^pu^hvu^hu^py$ , this also implies  $u^hvu^h = z_1az_2$ . Let  $\ell = (p - 1)(2h + 1)$ . We prove that  $w_1az_2u^\ell z_1 \in L_1$  and  $z_2u^\ell z_1aw_2 \in L_2$ . Since we have  $w_1 \in L_1$  and  $w_2 \in L_2$ , this yields that  $w_1az_2u^\ell z_1aw_2 \in L_1aL_2$  and that this membership is witnessed by two decompositions:  $w_1az_2u^\ell z_1 \cdot a \cdot w_2$  and  $w_1 \cdot a \cdot z_2u^\ell z_1aw_2$ . Hence, this contradicts the hypothesis that  $L_1aL_2$  is unambiguous, finishing the proof.

We only prove that  $w_1az_2u^\ell z_1 \in L_1$  (that  $z_2u^\ell z_1aw_2 \in L_2$  is symmetrical and left to the reader). Since  $w_1 = xu^pz_1$  and  $u^hvu^h = z_1az_2$ , we have,

$$w_1az_2u^\ell z_1 = xu^pu^hvu^hu^\ell z_1.$$

Thus, it remains to show that  $xu^pu^hvu^hu^\ell z_1 \in L_1$ . By definition,  $xu^pz_1 = w_1 \in L_1$ . Since  $p$  is the idempotent power of a finite monoid recognizing  $L_1$ , this yields  $xu^pu^{p(2h+1)}z_1 \in L_1$ . This can be reformulated as  $xu^pu^h u u^h u^\ell z_1 \in L_1$ . By definition,  $\eta(v) = \eta(u)$  and  $L_1$  satisfies (1) for  $k_1 \leq h$ . Hence, we obtain  $xu^pu^hvu^hu^\ell z_1 \in L_1$ , concluding the proof of the lemma. ■

We may now prove that  $xu^kv'u^ky \in L$ . In view of Lemma 4.27, there are two cases. First, assume that  $xu^kvu^h$  is a prefix of  $w_1$ . This yields  $z \in A^*$  such that  $w_1 = xu^kvu^hz$  and  $u^{k-h}y = zaw_2$ . Since  $w_1 \in L_1$ ,  $\eta(u) = \eta(v) = \eta(v')$  and  $L_1$  satisfies (1) for  $k_1 \leq h$ , it follows that,

$xu^k v' u^h z \in L_1$ . Thus,  $xu^k v' u^k y = xu^k v' u^h z a w_2 \in L_1 a L_2$  as desired. In the second case,  $u^h v u^k y$  is a suffix of  $w_2$ . This yields  $z \in A^*$  such that  $w_2 = zu^h v u^k y$  and  $xu^{k-h} = w_1 a z$ . Since  $w_2 \in L_2$ ,  $\eta(u) = \eta(v) = \eta(v')$  and  $L_2$  satisfies (1) for  $k_2 \leq h$ , it follows that,  $zu^h v' u^k y \in L_2$ . We obtain  $xu^k v' u^k y = w_1 a zu^h v' u^k y \in L_1 a L_2$ , as desired. ■

### 4.3 Alternated polynomial closure

We present a third and final operator. For every input class  $C$ , we define  $APol(C)$  as the least class containing  $C$  closed under *disjoint union*, *left deterministic marked concatenation* and *right deterministic marked concatenation*.

We now consider a technical restriction of  $APol$ , defined by further restricting the situations in which using marked concatenation is allowed. While less natural, it will be useful in proof arguments. Let  $C$  be a class of languages. Given two languages  $K, L \subseteq A^*$  and a letter  $a \in A$ , we say that the marked concatenation  $KaL$  is left (resp. right)  $C$ -deterministic when there exist  $K', L' \in C$  such that  $K \subseteq K', L \subseteq L'$  and  $K'aL'$  is left (resp. right) deterministic.

**REMARK 4.28.** Clearly,  $KaL$  being left (resp. right)  $C$ -deterministic implies that it is also left (resp. right) deterministic. The converse need not be true. For example, when  $C = ST = \{\emptyset, A^*\}$ , no nonempty marked concatenation can be left or right  $ST$ -deterministic.

The *weak alternated polynomial closure* of a class  $C$ , which we denote by  $WAPol(C)$ , is the least class containing  $C$  and closed under *disjoint union* and *left and right  $C$ -deterministic marked concatenation*. Clearly, we have  $WAPol(C) \subseteq APol(C)$ . Moreover, we shall prove in Section 7 that when  $C$  is a prevariety, the converse inclusion holds as well, so that  $WAPol(C) = APol(C)$ . This is useful to prove that  $APol(C)$  is included in another class: it suffices to prove that  $WAPol(C)$  is included in this class, which is often simpler.

While  $APol$  and  $WAPol$  are less prominent than  $UPol$ , they serve as key tools in the paper. This is because of the following theorem that we shall prove in Section 7.

**THEOREM 4.29.** *Let  $C$  be a prevariety. Then,  $UPol(C) = APol(C) = WAPol(C)$ .*

Recall that by Lemma 4.14, a left/right deterministic marked concatenation is necessarily unambiguous. Hence, the inclusions  $WAPol(C) \subseteq APol(C) \subseteq UPol(C)$  in Theorem 4.29 are immediate from the definitions. On the other hand, the proof argument for the inclusion  $UPol(C) \subseteq WAPol(C)$  is intertwined with that of the generic algebraic characterization of  $UPol(C)$ . More precisely, we prove that every language satisfying the characterization belongs to  $WAPol(C)$  (which also implies membership in  $UPol(C)$ ). This is why  $APol(C)$  and  $WAPol(C)$  are important. In practice, left/right deterministic marked concatenations are simpler to handle than unambiguous ones. By Theorem 4.29, they suffice to build *all* languages in  $UPol(C)$ .

## 5. Canonical relations associated to morphisms

Given a class  $C$  and a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid, we associate three distinct objects: a submonoid of  $M$  and two relations. They are central ingredients for formulating the algebraic characterizations of  $Pol(C)$  and  $UPol(C)$  that we present in Sections 6, 7 and 10. Our characterization algorithms rely on these notions for  $C$ : the  $C$ -kernel, the  $C$ -pair and the  $C$ -preorder relations.

### 5.1 $\mathcal{G}$ -kernels

Consider a class  $\mathcal{G}$  and a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid  $M$ . We associate to  $\alpha$  a subset  $N \subseteq M$  that we call the  $\mathcal{G}$ -kernel of  $\alpha$ . It consists of all elements  $s \in M$  such that  $\{\varepsilon\}$  is *not*  $\mathcal{G}$ -separable from  $\alpha^{-1}(s)$ . Additionally, we shall consider a slightly more restrictive notion: the *strict  $\mathcal{G}$ -kernel* of  $\alpha$  is the set  $S = N \cap \alpha(A^+)$ .

**REMARK 5.1.** While the definition makes sense for an arbitrary class  $\mathcal{G}$ , it is meant to be used in the special case where  $\mathcal{G}$  is a prevariety of group languages (hence the notation  $\mathcal{G}$ ).

**REMARK 5.2.** When  $\mathcal{G}$  is the class MOD of modulo languages, it can be shown that the MOD-kernel of a morphism corresponds to a standard notion: the stable monoid, defined in [37]. Given a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid, it can be verified that there exists a number  $d \geq 1$  such that  $\alpha(A^{2d}) = \alpha(A^d)$ . The least such number  $d \geq 1$  is called the stability index of  $\alpha$ . The stable semigroup of  $\alpha$  is  $S = \alpha(A^d)$  and the stable monoid of  $\alpha$  is  $N = \{1_M\} \cup S$ . One may verify that  $S$  and  $N$  are respectively the strict MOD-kernel and the MOD-kernel of  $\alpha$  (this follows from a simple analysis of MOD-separation).

Clearly, having a  $\mathcal{G}$ -separation algorithm in hand suffices to compute the  $\mathcal{G}$ -kernel of an input morphism  $\alpha$ . This yields the following lemma.

**LEMMA 5.3.** Let  $\mathcal{G}$  be a class with decidable separation. Given a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid as input, one can compute both the  $\mathcal{G}$ -kernel and the strict  $\mathcal{G}$ -kernel of  $\alpha$ .

We now characterize  $\mathcal{G}$ -kernels in terms of  $\mathcal{G}$ -morphisms when  $\mathcal{G}$  is a prevariety.

**LEMMA 5.4.** Let  $\mathcal{G}$  be a prevariety, let  $\alpha : A^* \rightarrow M$  be a morphism into a finite monoid and let  $N$  be the  $\mathcal{G}$ -kernel of  $\alpha$ . The following properties hold:

- Let  $\eta : A^* \rightarrow G$  be a  $\mathcal{G}$ -morphism. For every  $s \in N$ , there exists  $w \in A^*$  such that  $\alpha(w) = s$  and  $\eta(w) = 1_G$ .
- There exists a  $\mathcal{G}$ -morphism  $\eta : A^* \rightarrow G$  such that for every  $w \in A^*$ , if  $\eta(w) = 1_G$ , then  $\alpha(w) \in N$ .

**PROOF.** For the first assertion, let  $\eta : A^* \rightarrow G$  be a  $\mathcal{G}$ -morphism and consider  $s \in N$ . By hypothesis,  $\eta^{-1}(1_G) \in \mathcal{G}$  and  $\{\varepsilon\} \subseteq \eta^{-1}(1_G)$ . Since  $s \in N$  and  $N$  is the  $\mathcal{G}$ -kernel of  $\alpha$ , the language  $\eta^{-1}(1_G) \in \mathcal{G}$  cannot separate  $\alpha^{-1}(s)$  from  $\{\varepsilon\}$ . Since it includes  $\{\varepsilon\}$ , we have  $\alpha^{-1}(s) \cap \eta^{-1}(1_G) \neq \emptyset$ . This yields  $w \in A^*$  such that  $\alpha(w) = s$  and  $\eta(w) = 1_G$ .

We turn to the second assertion. By definition of the  $\mathcal{G}$ -kernel  $N$  of  $\alpha$ , we know that for every  $t \in M \setminus N$ , there exists  $K_t \in \mathcal{G}$  such that  $\varepsilon \in K_t$  and  $\alpha^{-1}(t) \cap K_t = \emptyset$ . Proposition 2.13 yields a  $\mathcal{G}$ -morphism  $\eta : A^* \rightarrow G$  recognizing all languages  $K_t$  for  $t \in M \setminus N$ . It remains to prove that given  $w \in A^*$  such that  $\eta(w) = 1_G$ , we have  $\alpha(w) \in N$ . Assume that  $\eta(w) = 1_G = \eta(\varepsilon)$ . Since  $\eta$  recognizes  $K_t$  for every  $t \in M \setminus N$  and  $\varepsilon \in K_t$  by definition, we get  $w \in K_t$  for every  $t \in M \setminus N$ . Since  $\alpha^{-1}(t) \cap K_t = \emptyset$ , we deduce that  $\alpha(w) \neq t$  for every  $t \in M \setminus N$ . This means that  $\alpha(w) \in N$ , completing the proof. ■

An important property is that the  $\mathcal{G}$ -kernel  $N$  of a morphism  $\alpha : A^* \rightarrow M$  is a submonoid of  $M$ . Observe that since the *strict*  $\mathcal{G}$ -kernel of  $\alpha$  is  $S = N \cap \alpha(A^+)$  by definition, this also implies that  $S$  is a subsemigroup of  $M$ . We state these properties in the following fact, which we shall use implicitly from now on.

**FACT 5.5.** Let  $\mathcal{G}$  be a prevariety and  $\alpha : A^* \rightarrow M$  a morphism into a finite monoid. The  $\mathcal{G}$ -kernel of  $\alpha$  is a submonoid of  $M$ . Its strict  $\mathcal{G}$ -kernel is a subsemigroup of  $M$ .

**PROOF.** We write  $N$  for the  $\mathcal{G}$ -kernel of  $\alpha$ . Lemma 5.4 yields a  $\mathcal{G}$ -morphism  $\eta : A^* \rightarrow G$  such that for every  $w \in A^*$ , if  $\eta(w) = 1_G$ , then  $\alpha(w) \in N$ . Since  $\eta(\varepsilon) = 1_G$ , it follows that  $1_M = \alpha(\varepsilon) \in N$ . Moreover, if  $s, t \in N$ , then the first assertion in Lemma 5.4 yields  $u, v \in A^*$  such that  $\eta(u) = \eta(v) = 1_G$ ,  $\alpha(u) = s$  and  $\alpha(v) = t$ . Hence,  $\eta(uv) = 1_G$  and we get  $st = \alpha(uv) \in N$  by definition of  $\eta$ . ■

We conclude the presentation with a lemma, specific to the case where  $\mathcal{G}$  is a prevariety of group languages (which is the only case that we shall consider in practice).

**FACT 5.6.** Let  $\mathcal{G}$  be a prevariety of group languages and let  $\alpha : A^* \rightarrow M$  be a surjective morphism into a finite monoid. Then, all idempotents in  $E(M)$  belongs to the  $\mathcal{G}$ -kernel of  $\alpha$ .

**PROOF.** Let  $N$  be the  $\mathcal{G}$ -kernel of  $\alpha$  and let  $e \in E(M)$ . We prove that  $e \in N$ . By Lemma 5.4, there exists a  $\mathcal{G}$ -morphism  $\eta : A^* \rightarrow G$  such that for every  $w \in A^*$ , if  $\eta(w) = 1_G$ , then  $\alpha(w) \in N$ . By Lemma 2.14, we know that  $G$  is a group. Since  $\alpha$  is surjective, there exists  $w \in A^*$  such that  $\alpha(w) = e$ . Let  $k = \omega(G)$ . Since  $G$  is a group, we get  $\eta(w^k) = 1_G$  by definition of  $k$ . Moreover, since  $e$  is idempotent, we have  $\alpha(w^k) = e$ . Hence,  $e \in N$  by definition of  $\eta$ . ■

## 5.2 $C$ -pairs

Consider a class  $C$  and a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid. We define a relation on  $M$ : the  $C$ -pairs (for  $\alpha$ ). Let  $(s, t) \in M^2$ . We say that,

$$(s, t) \text{ is a } C\text{-pair (for } \alpha) \text{ if and only if } \alpha^{-1}(s) \text{ is not } C\text{-separable from } \alpha^{-1}(t). \quad (2)$$

**REMARK 5.7.** While we often make this implicit, being a  $C$ -pair depends on  $\alpha$ .

By definition, the set of  $C$ -pairs for  $\alpha$  is finite: it is a subset of  $M^2$ . Moreover, having a  $C$ -separation algorithm in hand is clearly enough to compute all  $C$ -pairs associated to an input morphism  $\alpha$ . We state this simple, yet crucial property in the following lemma.

**LEMMA 5.8.** Let  $C$  be a class of languages with decidable separation. Then, given a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid as input, one can compute all  $C$ -pairs for  $\alpha$ .

We complete the definition with some properties of  $C$ -pairs. A simple and useful one is that the  $C$ -pair relation is reflexive provided that the morphism  $\alpha$  is surjective (which is always the case in practice). Moreover, it is symmetric when  $C$  is closed under complement. On the other hand, the  $C$ -pair relation is *not* transitive in general (see Example 5.10 below).

**LEMMA 5.9.** Consider a class  $C$  and a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid. The following properties hold:

- If  $\alpha$  is surjective, the  $C$ -pair relation is reflexive: for every  $s \in M$ ,  $(s, s)$  is a  $C$ -pair.
- If  $C$  is closed under complement, the  $C$ -pair relation is symmetric: for every  $C$ -pair  $(s, t) \in M^2$ ,  $(t, s)$  is a  $C$ -pair as well.

**PROOF.** For first assertion, assume that  $\alpha$  is surjective. Given  $s \in M$ , we have  $\alpha^{-1}(s) \neq \emptyset$  since  $\alpha$  is surjective. Hence,  $\alpha^{-1}(s) \cap \alpha^{-1}(s) \neq \emptyset$ , which implies that  $\alpha^{-1}(s)$  is not  $C$ -separable from  $\alpha^{-1}(s)$ . This exactly says that  $(s, s)$  is a  $C$ -pair. For the second assertion, assume that  $C$  is closed under complement and consider a  $C$ -pair  $(s, t) \in M^2$ . It follows that  $\alpha^{-1}(s)$  is not  $C$ -separable from  $\alpha^{-1}(t)$ . Since  $C$  is closed under complement, this implies that  $\alpha^{-1}(t)$  is not  $C$ -separable from  $\alpha^{-1}(s)$ . Therefore,  $(t, s)$  is a  $C$ -pair. ■

**EXAMPLE 5.10.** The  $C$ -pair relation is not transitive in general. Indeed, let  $C = \text{AT}$  and let  $M$  be the monoid  $\{1, a, b, 0\}$  where 1 acts as an identity element, 0 as an absorbing element, and the rest of the multiplication is given by  $aa = ab = ba = bb = 0$ . Let  $A = \{a, b\}$  and  $\alpha : A^* \rightarrow M$  be the morphism defined by  $\alpha(a) = a$  and  $\alpha(b) = b$ . We have  $\alpha^{-1}(a) = \{a\}$ . Therefore, any language of AT containing  $\alpha^{-1}(a)$  also contains  $a^+$ , and intersects  $\alpha^{-1}(0)$  (which is the set of words of length at least 2). Hence,  $(a, 0)$  is an AT-pair. Likewise,  $(0, b)$  is an AT-pair. However,  $(a, b)$  is not an AT-pair, since the language  $a^+ \in \text{AT}$  separates  $\alpha^{-1}(a) = \{a\}$  from  $\alpha^{-1}(b) = \{b\}$ .

We now provide a useful characterization of  $C$ -pairs via  $C$ -morphisms in the special case where  $C$  is a positive prevariety, which is the only case that we consider in practice. It is similar to Lemma 5.4.

**LEMMA 5.11.** Let  $C$  be a positive prevariety and let  $\alpha : A^* \rightarrow M$  be a morphism into a finite monoid. The two following properties hold:

- For every  $C$ -morphism  $\eta : A^* \rightarrow (N, \leq)$  and every  $C$ -pair  $(s, t) \in M^2$  for  $\alpha$ , there exist  $u, v \in A^*$  such that  $\eta(u) \leq \eta(v)$ ,  $\alpha(u) = s$  and  $\alpha(v) = t$ .
- There exists a  $C$ -morphism  $\eta : A^* \rightarrow (N, \leq)$  such that for all  $u, v \in A^*$ , if  $\eta(u) \leq \eta(v)$ , then  $(\alpha(u), \alpha(v))$  is a  $C$ -pair for  $\alpha$ .

**PROOF.** Let us start with the first assertion. Let  $\eta : A^* \rightarrow (N, \leq)$  be a  $C$ -morphism and let  $(s, t) \in M^2$  be a  $C$ -pair for  $\alpha$ . Let  $F \subseteq N$  be the set of all elements  $r \in N$  such that  $\eta(u) \leq r$  for some  $u \in \alpha^{-1}(s)$ . By definition,  $F$  is an upper set for the ordering  $\leq$  on  $N$ . Hence, we have  $\eta^{-1}(F) \in C$  since  $\eta$  is a  $C$ -morphism. Moreover, it is immediate from the definition of  $F$  that  $\alpha^{-1}(s) \subseteq \eta^{-1}(F)$ . Therefore, since  $(s, t)$  is a  $C$ -pair (which means that  $\alpha^{-1}(s)$  cannot be separated from  $\alpha^{-1}(t)$  using a language in  $C$ ), it follows that  $\eta^{-1}(F) \cap \alpha^{-1}(t) \neq \emptyset$ . This yields  $v \in A^*$  such that  $\eta(v) \in F$  and  $\alpha(v) = t$ . Finally, since  $v \in F$ , the definition of  $F$  yields  $u \in A^*$  such that  $\eta(u) \leq \eta(v)$  and  $\alpha(u) = s$ , concluding the proof of the first assertion.

We turn to the second assertion. Let  $P \subseteq M^2$  be the set of all pairs  $(s, t) \in M^2$  which are *not*  $C$ -pairs. For every  $(s, t) \in P$ , there exists  $K_{s,t} \in C$  separating  $\alpha^{-1}(s)$  from  $\alpha^{-1}(t)$ . Proposition 2.13 yields a  $C$ -morphism  $\eta : A^* \rightarrow (N, \leq)$  such that every language  $K_{s,t}$  for  $(s, t) \in P$  is recognized by  $\eta$ . It remains to prove that for every  $u, v \in A^*$ , if  $\eta(u) \leq \eta(v)$ , then  $(\alpha(u), \alpha(v))$  is a  $C$ -pair. We prove the contrapositive. Assuming that  $(\alpha(u), \alpha(v))$  is *not* a  $C$ -pair, we show that  $\eta(u) \not\leq \eta(v)$ . By hypothesis,  $(\alpha(u), \alpha(v)) \in P$ , which means that  $K_{\alpha(u), \alpha(v)} \in C$  is defined and separates  $\alpha^{-1}(\alpha(u))$  from  $\alpha^{-1}(\alpha(v))$ . Thus,  $u \in K_{\alpha(u), \alpha(v)}$  and  $v \notin K_{\alpha(u), \alpha(v)}$ . Since  $K_{\alpha(u), \alpha(v)}$  is recognized by  $\eta$ , this implies  $\eta(u) \not\leq \eta(v)$ . ■

We now prove that when  $C$  is a positive prevariety of regular languages (which is the only case that we shall consider), the  $C$ -pair relation is compatible with multiplication. This result is similar to Fact 5.5 for  $\mathcal{G}$ -kernels.

**LEMMA 5.12.** Let  $C$  be a positive prevariety and let  $\alpha : A^* \rightarrow M$  be a morphism into a finite monoid. If  $(s_1, t_1), (s_2, t_2) \in M^2$  are  $C$ -pairs, then  $(s_1s_2, t_1t_2)$  is a  $C$ -pair as well.

**PROOF.** Lemma 5.11 yields a  $C$ -morphism  $\eta : A^* \rightarrow (N, \leq)$  such that for all  $u, v \in A^*$ , if  $\eta(u) \leq \eta(v)$ , then  $(\alpha(u), \alpha(v))$  is a  $C$ -pair. Let  $(s_1, t_1), (s_2, t_2) \in M^2$  be  $C$ -pairs. Since  $\eta$  is a  $C$ -morphism, it follows from Lemma 5.11 that there exist  $u_i, v_i \in A^*$  for  $i = 1, 2$  such that  $\eta(u_i) \leq \eta(v_i)$ ,  $\alpha(u_i) = s_i$  and  $\alpha(v_i) = t_i$ . Clearly, it follows that  $\eta(u_1u_2) \leq \eta(v_1v_2)$ ,  $\alpha(u_1u_2) = s_1s_2$  and  $\alpha(v_1v_2) = t_1t_2$ . Hence,  $(s_1s_2, t_1t_2)$  is a  $C$ -pair by definition of  $\eta$ . ■

We conclude the presentation with a result connecting  $\mathcal{G}$ -pairs with the  $\mathcal{G}$ -kernel, when  $\mathcal{G}$  is a prevariety of group languages.

**LEMMA 5.13.** Let  $\mathcal{G}$  be a prevariety of group languages and let  $\alpha : A^* \rightarrow M$  be a surjective morphism into a finite monoid. Let  $e \in E(M)$  and  $s, t \in M$  such that  $(e, s)$  is a  $\mathcal{G}$ -pair. Then,  $es$  and  $t(est)^{2\omega-1}$  both belong to the  $\mathcal{G}$ -kernel of  $\alpha$ .

**PROOF.** Let  $N$  be the  $\mathcal{G}$ -kernel of  $\alpha$ . Lemma 5.4 yields a  $\mathcal{G}$ -morphism  $\eta : A^* \rightarrow G$  such that for every  $w \in A^*$ , if  $\eta(w) = 1_G$ , then  $\alpha(w) \in N$ . By Lemma 2.14,  $G$  is a group. We first show that  $es \in N$ . Since  $(e, s)$  is a  $\mathcal{G}$ -pair, Lemma 5.11 yields  $u, v \in A^*$  such that  $\eta(u) = \eta(v)$ ,  $\alpha(u) = e$  and  $\alpha(v) = s$ . Let  $p = \omega(G) \times \omega(M)$ . Since  $\eta(u) = \eta(v)$ , we have  $\eta(u^{p-1}v) = \eta(u^p) = 1_G$  (recall that  $G$  is a group). Hence, by definition of  $\eta$ , we have  $\alpha(u^{p-1}v) \in N$ . Since  $e$  is an idempotent, this yields  $es \in N$ , as desired.

We now prove that  $t(est)^{2\omega-1} \in N$ . Let  $w = u^{p-1}v$ . We know that  $\eta(w) = 1_G$  and  $\alpha(w) = es$ . Since  $\alpha$  is surjective, we get  $x \in A^*$  such that  $\alpha(x) = t$ . Since  $\eta(w) = 1_G$  and  $p$  is a multiple of  $\omega(G)$ , we have  $\eta(x(wx)^{2p-1}) = \eta(x^{2p}) = 1_G$ . Moreover, since  $p$  is a multiple of  $\omega(M)$ , we have  $\alpha(x(wx)^{2p-1}) = t(est)^{2\omega-1}$ . By definition of  $\eta$ , we get  $t(est)^{2\omega-1} \in N$ . ■

### 5.3 Canonical $C$ -preorder and $C$ -equivalence

Consider a *lattice*  $C$ . Moreover, let  $\alpha : A^* \rightarrow M$  be a morphism into a finite monoid. We define two relations on  $M$ : a preorder “ $\leq_{C,\alpha}$ ” which we call the *canonical  $C$ -preorder* of  $\alpha$  and an equivalence “ $\sim_{C,\alpha}$ ” which we call the *canonical  $C$ -equivalence* of  $\alpha$ . Consider a pair  $(s, t) \in M^2$ . We define,

$$s \leq_{C,\alpha} t \quad \text{if and only if} \quad s \in F \Rightarrow t \in F \quad \text{for all } F \subseteq M \text{ such that } \alpha^{-1}(F) \in C. \quad (3)$$

$$s \sim_{C,\alpha} t \quad \text{if and only if} \quad s \in F \Leftrightarrow t \in F \quad \text{for all } F \subseteq M \text{ such that } \alpha^{-1}(F) \in C. \quad (4)$$

It is immediate by definition that  $\leq_{C,\alpha}$  is indeed a preorder on  $M$ . Moreover,  $\sim_{C,\alpha}$  is exactly the equivalence induced by  $\leq_{C,\alpha}$ : for every  $s, t \in M$ , we have  $s \sim_{C,\alpha} t$  if and only if  $s \leq_{C,\alpha} t$  and  $t \leq_{C,\alpha} s$ . From now on, for the sake of avoiding clutter, we shall abuse terminology when the morphism  $\alpha$  is understood and we write  $\leq_C$  for  $\leq_{C,\alpha}$  and  $\sim_C$  for  $\sim_{C,\alpha}$ . Additionally, for every element  $s \in M$ , we write  $[s]_C \in M/\sim_C$  for the  $\sim_C$ -class of  $s$ . Finally, note that since  $\sim_C$  is the equivalence induced by  $\leq_C$ , the preorder  $\leq_C$  also induces an order on the quotient set  $M/\sim_C$ , which we write  $\leq_C$ : for every  $s, t \in M$  we have  $[s]_C \leq_C [t]_C$  if and only if  $s \leq_C t$ .

Observe that an immediate key property of these relations is that having an algorithm for  $C$ -membership suffices to compute both  $\leq_C$  and  $\sim_C$ . Indeed, with such a procedure in hand, it is possible to compute all subsets  $F \subseteq M$  such that  $\alpha^{-1}(F) \in C$ . One may then decide whether  $s \leq_C t$  for some  $s, t \in M$  by checking whether the implication  $s \in F \Rightarrow t \in F$  holds for every such subset  $F$ . We state this in the following lemma.

**LEMMA 5.14.** Let  $C$  be a class of languages with decidable membership. Given a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid as input, one can compute the relations  $\leq_C$  and  $\sim_C$  on  $M$ .

**Properties.** First, we show that given a morphism  $\alpha : A^* \rightarrow M$ , we are able to characterize the languages simultaneously recognized by  $\alpha$  and belonging to  $C$  as those which are the inverse image of an *upper set* for the preorder  $\leq_C$ . Recall that by definition,  $F \subseteq M$  is an upper set for  $\leq_C$  if and only if for every  $s \in F$  and  $t \in M$  such that  $s \leq_C t$ , we have  $t \in F$ .

**LEMMA 5.15.** Consider a lattice  $C$  and a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid. For every  $F \subseteq M$ , we have  $\alpha^{-1}(F) \in C$  if and only if  $F$  is an upper set for  $\leq_C$ .

**PROOF.** We fix  $F \subseteq M$  for the proof. Assume first that  $\alpha^{-1}(F) \in C$  we prove that  $F$  is an upper set for  $\leq_C$ . Let  $s \in F$  and  $t \in M$  such that  $s \leq_C t$ . By definition of  $\leq_C$  and since  $\alpha^{-1}(F) \in C$ , this implies that  $t \in F$  as desired. Conversely assume that  $F$  is an upper set for  $\leq_C$ . We prove that  $\alpha^{-1}(F) \in C$ . Consider  $s \in F$ . By definition, we know that for every element  $r \notin F$ , we have  $s \not\leq_C r$ . Hence by definition of  $\leq_C$ , there exists a set  $F_{s,r} \subseteq M$  such that  $\alpha^{-1}(F_{s,r}) \in C$ ,  $s \in F_{s,r}$  and  $r \notin F_{s,r}$ . It is now immediate that,

$$F = \bigcup_{s \in F} \left( \bigcap_{r \notin F} F_{s,r} \right).$$

Since inverse image commutes with Boolean operations, we obtain,

$$\alpha^{-1}(F) = \bigcup_{s \in F} \left( \bigcap_{r \notin F} \alpha^{-1}(F_{s,r}) \right).$$

We conclude that  $\alpha^{-1}(F) \in C$ , since  $C$  is a lattice. ■

We turn to a key property of the preorder  $\leq_C$ . We connect it to the  $C$ -pair relation associated to every morphism  $\alpha : A^* \rightarrow M$ . Specifically, we show that  $\leq_C$  is the reflexive transitive closure of the  $C$ -pair relation.

**LEMMA 5.16.** Consider a lattice  $C$  and a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid. Then, the relation  $\leq_C$  on  $M$  is reflexive transitive closure of the  $C$ -pair relation.

**PROOF.** We first prove that  $\leq_C$  contains the reflexive transitive closure of the  $C$ -pair relation. Since  $\leq_C$  is a preorder, it suffices to show that for every  $C$ -pair  $(q, r) \in M^2$ , we have  $q \leq_C r$ . Hence, we fix  $F \subseteq M$  such that  $\alpha^{-1}(F) \in C$  and  $q \in F$ . We have to show that  $r \in F$ . Clearly,  $q \in F$  implies that  $\alpha^{-1}(q) \subseteq \alpha^{-1}(F)$ . Since  $(q, r)$  is a  $C$ -pair and  $\alpha^{-1}(F) \in C$ , we know that  $\alpha^{-1}(r) \cap \alpha^{-1}(F) \neq \emptyset$ , whence  $r \in F$ , as desired.

We turn to the converse implication. Let  $s, t \in M$  be such that  $s \leq_C t$ . Moreover, let  $F \subseteq M$  be the least subset of  $M$  containing  $s$  and such that for every  $C$ -pair  $(q, r) \in M^2$ , if  $q \in F$ , then  $r \in F$  as well. We have to show that  $t \in F$ . Since  $s \leq_C t$ , it suffices to show that  $\alpha^{-1}(F) \in C$  by definition of  $\leq_C$ . For every  $q \in F$ , we may build a language  $H_q \in C$  such that

$\alpha^{-1}(q) \subseteq H_q \subseteq \alpha^{-1}(F)$ . Indeed, for any  $r \notin F$ , we know that  $(q, r)$  is not a  $C$ -pair by definition of  $F$ . Therefore, there exists  $H_{q,r} \in C$  separating  $\alpha^{-1}(q)$  from  $\alpha^{-1}(r)$ . We let,

$$H_q = \bigcap_{r \notin F} H_{q,r}.$$

Clearly  $H_q \in C$  since  $C$  is a lattice. It now suffices to observe that,

$$\alpha^{-1}(F) = \bigcup_{q \in F} \alpha^{-1}(q) \subseteq \bigcup_{q \in F} H_q \subseteq \alpha^{-1}(F).$$

Therefore,  $\alpha^{-1}(F) = \bigcup_{q \in F} H_q$  belongs to  $C$  since  $C$  is lattice. This concludes the proof. ■

We now consider the case where  $C$  is a *Boolean algebra*. We show that in this case the preorder  $\leq_C$  coincides with the equivalence  $\sim_C$  for every morphism  $\alpha : A^* \rightarrow M$ .

**LEMMA 5.17.** Consider a Boolean algebra  $C$  and a morphism  $\alpha : A^* \rightarrow M$  into a finite monoid. For every  $s, t \in M$ , we have  $s \leq_C t$  if and only if  $s \sim_C t$ .

**PROOF.** We fix  $s, t \in M$  for the proof. It is immediate by definition that if  $s \sim_C t$ , then we have  $s \leq_C t$ . We prove the converse implication. Assume that  $s \leq_C t$ . For  $F \subseteq M$  such that  $\alpha^{-1}(F) \in C$ , we need to show that  $s \in F \Leftrightarrow t \in F$ . The implication  $s \in F \Rightarrow t \in F$  is immediate from  $s \leq_C t$ . For the converse direction, since  $\alpha^{-1}(F) \in C$  and  $C$  is a Boolean algebra, we have  $\alpha^{-1}(M \setminus F) = A^* \setminus \alpha^{-1}(F) \in C$ . Hence,  $s \leq_C t$  yields  $s \in M \setminus F \Rightarrow t \in M \setminus F$ . The contrapositive exactly says that  $t \in F \Rightarrow s \in F$ , which concludes the proof. ■

Finally, we consider the particular case where  $C$  is additionally closed under quotients (i.e., when  $C$  is a positive prevariety). We show that in this case  $\leq_C$  and  $\sim_C$  are compatible with the multiplication of  $M$ , provided that the morphism  $\alpha : A^* \rightarrow M$  is surjective.

**LEMMA 5.18.** Let  $C$  be a positive prevariety and let  $\alpha : A^* \rightarrow M$  be a surjective morphism into a finite monoid. The two following properties hold for every  $s_1, t_1, s_2, t_2 \in M$ :

- if  $s_1 \leq_C t_1$  and  $s_2 \leq_C t_2$ , then  $s_1 s_2 \leq_C t_1 t_2$ .
- if  $s_1 \sim_C t_1$  and  $s_2 \sim_C t_2$ , then  $s_1 s_2 \sim_C t_1 t_2$ .

**PROOF.** Recall that by definition,  $s \sim_C t$  if and only if  $s \leq_C t$  and  $t \leq_C s$  for every  $s, t \in M$ . Hence, it suffices to prove the first assertion as the second is an immediate consequence. We fix  $s_1, t_1, s_2, t_2 \in M$  such that  $s_1 \leq_C t_1$  and  $s_2 \leq_C t_2$ . We prove that  $s_1 s_2 \leq_C t_1 t_2$ . Let  $F \subseteq M$  such that  $\alpha^{-1}(F) \in C$  and  $s_1 s_2 \in F$ . We show that  $t_1 t_2 \in F$ . Let  $u, v \in A^*$  be two words such that  $\alpha(u) = s_1$  and  $\alpha(v) = t_2$  (this is where we use the hypothesis that  $\alpha$  is surjective). Since  $s_1 s_2 \in F$ , we have  $s_2 \in s_1^{-1}F$ . Moreover,  $\alpha^{-1}(s_1^{-1}F) = u^{-1}\alpha^{-1}(F)$  belongs to  $C$  by closure under quotients. Since  $s_2 \leq_C t_2$ , we get  $t_2 \in s_1^{-1}F$ . It follows that  $s_1 t_2 \in F$ , which means that  $s_1 \in Ft_2^{-1}$ . Moreover,  $\alpha^{-1}(Ft_2^{-1}) = \alpha^{-1}(F)v^{-1}$  belongs to  $C$  by closure under quotients. Since  $s_1 \leq_C t_1$ , we get  $t_1 \in Ft_2^{-1}$ . Altogether, we obtain  $t_1 t_2 \in F$ , which concludes the proof. ■

**Connection with  $C$ -morphisms.** When  $C$  is a positive prevariety, it follows from Lemma 5.18 that the canonical  $C$ -preorder  $\leq_C$  of a surjective morphism  $\alpha : A^* \rightarrow M$  is a *precongruence* on the monoid  $M$  and the induced equivalence  $\sim_C$  is a *congruence*. Consequently, the pair  $(M/\sim_C, \leq_C)$  is an ordered monoid (recall that  $\leq_C$  denotes the order induced by  $\leq_C$  on the quotient set  $M/\sim_C$ ). Moreover, the map  $[\cdot]_C : M \rightarrow M/\sim_C$  (which associates its  $\sim_C$ -class to every element in  $M$ ) is a morphism. In the following lemma, we prove that the morphism  $[\cdot]_C \circ \alpha : A^* \rightarrow (M/\sim_C, \leq_C)$  is a  $C$ -morphism.

**LEMMA 5.19.** Let  $C$  be a positive prevariety and let  $\alpha : A^* \rightarrow M$  be a surjective morphism into a finite monoid. The languages recognized by  $[\cdot]_C \circ \alpha$  are exactly those in  $C$  which are recognized by  $\alpha$ . In particular,  $[\cdot]_C \circ \alpha : A^* \rightarrow (M/\sim_C, \leq_C)$  is a  $C$ -morphism.

**PROOF.** It is immediate from the definitions that the languages recognized by  $[\cdot]_C \circ \alpha$  are exactly those of the form  $\alpha^{-1}(F)$  where  $F \subseteq M$  is an upper set for  $\leq_C$ . Hence, Lemma 5.15 implies that these are exactly the languages in  $C$  which are recognized by  $\alpha$ , concluding the proof. ■

## 6. Algebraic characterizations for polynomial closure

We present an algebraic characterization for the classes built with polynomial closure. The statements are based on Proposition 2.12: given a class  $\mathcal{D}$ , rather than directly characterizing the languages in  $\mathcal{D}$ , we characterize the  $\mathcal{D}$ -morphisms. Presenting the results in this form is more convenient for reusing them later. We first consider the classes  $Pol(C)$  when  $C$  is a positive prevariety. We recall their algebraic characterization proved in [26]. Then, we use it to prove generic characterizations of the classes  $Pol(C) \cap coPol(C)$  and  $Pol(BPol(C)) \cap coPol(BPol(C))$  when  $C$  is a positive prevariety.

### 6.1 Polynomial closure

We first present the generic characterization of the  $Pol(C)$ -morphisms proved in [26]. The statement is based on the  $C$ -pair relation defined in Section 5. Since classes of the form  $Pol(C)$  are *not* closed under complement in general, it is important to consider morphisms into finite *ordered* monoids here.

**THEOREM 6.1 ([26]).** Let  $C$  be a positive prevariety and let  $\alpha : A^* \rightarrow (M, \leq)$  be a surjective morphism into a finite ordered monoid. The following properties are equivalent:

1.  $\alpha$  is a  $Pol(C)$ -morphism.
2.  $\alpha$  satisfies the following property:

$$s^{\omega+1} \leq s^{\omega}ts^{\omega} \quad \text{for every } C\text{-pair } (s, t) \in M^2. \quad (5)$$

Given a positive prevariety  $C$ , it follows from Lemma 5.8 that the  $C$ -pairs associated to a morphism can be computed provided that  $C$ -separation is decidable. Hence, Theorem 6.1 implies that in this case, one can decide whether an input morphism  $\alpha : A^* \rightarrow (M, \leq)$  is a  $Pol(C)$ -morphism. In view of Proposition 2.12, this implies that  $Pol(C)$ -membership is decidable. Indeed, given as input a regular language  $L$ , one can compute its syntactic morphism and since  $Pol(C)$  is a positive prevariety by Theorem 4.3, the proposition yields that  $L \in Pol(C)$  if and only if this syntactic morphism is a  $Pol(C)$ -morphism. Altogether, we obtain the following corollary.

**COROLLARY 6.2.** Let  $C$  be a positive prevariety with decidable separation. Then  $Pol(C)$ -membership is decidable.

**REMARK 6.3.** It is also known that  $Pol(C)$ -covering and  $Pol(C)$ -separation are decidable when  $C$  is a finite prevariety [22]. We do not detail this result as we shall not need it.

## 6.2 Intersected polynomial closure

With the generic characterization of  $Pol(C)$  in hand, it is straightforward to deduce another generic characterization for  $Pol(C) \cap coPol(C)$ . In particular, it shows that membership for  $Pol(C) \cap coPol(C)$  boils down to separation for  $C$  (when  $C$  is a positive prevariety). However, it turns out that this result can be strengthened: for any positive prevariety  $C$ , membership for  $Pol(C) \cap coPol(C)$  reduces to *membership* for  $C$ . This is a simple consequence of Theorem 6.1. It was first observed by Almeida, Bartonová, Klíma and Kunc [2]. Indeed, we have the following theorem. It is based on the canonical preorder  $\leq_C$  associated to a morphism  $\alpha$  (note that since classes of the form  $Pol(C) \cap coPol(C)$  are prevarieties, it suffices to consider morphisms into unordered monoids here, see Remark 2.11).

**THEOREM 6.4.** Let  $C$  be a positive prevariety and let  $\alpha : A^* \rightarrow M$  be a surjective morphism into a finite monoid. The following properties are equivalent:

1.  $\alpha$  is a  $(Pol(C) \cap coPol(C))$ -morphism.
2.  $\alpha$  satisfies the following property:

$$s^{\omega+1} = s^{\omega} t s^{\omega} \quad \text{for every } C\text{-pair } (s, t) \in M^2. \quad (6)$$

3.  $\alpha$  satisfies the following property:

$$s^{\omega+1} = s^{\omega} t s^{\omega} \quad \text{for every } s, t \in M \text{ such that } s \leq_C t. \quad (7)$$

**PROOF.** We first prove (1)  $\Leftrightarrow$  (2). Recall that we view  $\alpha$  as the morphism  $\alpha : A^* \rightarrow (M, =)$  into the ordered monoid  $(M, =)$ . Therefore, we get from Theorem 6.1 that  $\alpha$  is a  $Pol(C)$ -morphism if

and only if (6) is satisfied. It remains to prove that  $\alpha$  is a  $(Pol(C) \cap coPol(C))$ -morphism if and only if it is a  $Pol(C)$ -morphism. The left to right implication is immediate. For the converse one, assume that  $\alpha$  is a  $Pol(C)$ -morphism and let  $F \subseteq M$ . We show that  $\alpha^{-1}(F) \in Pol(C) \cap coPol(C)$ . By hypothesis, we know that  $\alpha^{-1}(F) \in Pol(C)$  and  $A^* \setminus \alpha^{-1}(F) = \alpha^{-1}(M \setminus F) \in Pol(C)$ . Hence,  $\alpha^{-1}(F) \in Pol(C) \cap coPol(C)$ , as desired.

The implication (3)  $\Rightarrow$  (2) is immediate from Lemma 5.16, which entails that for every  $C$ -pair  $(s, t) \in M^2$ , we have  $s \leq_C t$ . It remains to show the implication (2)  $\Rightarrow$  (3). Assuming that  $\alpha$  satisfies (6), we show it satisfies (7) as well. Let  $s, t \in M^2$  such that  $s \leq_C t$ . We show that  $s^{\omega+1} = s^\omega t s^\omega$ . Lemma 5.16 yields  $k \in \mathbb{N}$  and  $r_0, \dots, r_k \in M$  such that  $r_0 = s, r_k = t$  and  $(r_i, r_{i+1})$  is a  $C$ -pair for all  $i < k$ . Using induction, we show that for all  $i \geq 1$ ,  $s^{\omega+1} = s^\omega r_i s^\omega$ . The case  $i = k$  then yields the desired result, as  $t = r_k$ . When  $i = 0$ , it is immediate that  $s^{\omega+1} = s^\omega r_0 s^\omega$  since  $r_0 = s$ . We now assume that  $i \geq 1$ . Using induction, we get that  $s^{\omega+1} = s^\omega r_{i-1} s^\omega$ . Therefore, we obtain  $s^\omega = (s^{\omega+1})^\omega = (s^\omega r_{i-1} s^\omega)^\omega$ . Since  $(r_{i-1}, r_i)$  is a  $C$ -pair, it follows from Lemma 5.12 that  $(s^\omega r_{i-1} s^\omega, s^\omega r_i s^\omega)$  is a  $C$ -pair as well. Thus, it follows from (6) that,

$$(s^\omega r_{i-1} s^\omega)^{\omega+1} = (s^\omega r_{i-1} s^\omega)^\omega s^\omega r_i s^\omega (s^\omega r_{i-1} s^\omega)^\omega.$$

Since  $s^{\omega+1} = s^\omega r_{i-1} s^\omega$  and  $s^\omega = (s^\omega r_{i-1} s^\omega)^\omega$ , this yields,

$$s^{\omega+1} = (s^{\omega+1})^{\omega+1} = s^\omega s^\omega r_i s^\omega s^\omega = s^\omega r_i s^\omega.$$

This concludes the proof. ■

Theorem 6.4 has the announced consequence: if  $C$  is a positive prevariety with decidable membership, then  $(Pol(C) \cap coPol(C))$ -membership is decidable as well. Indeed, given as input a regular language  $L$ , we can compute its syntactic morphism  $\alpha : A^* \rightarrow M$  together with the relation  $\leq_C$  on  $M$  (this is possible by Lemma 5.14 since  $C$ -membership is decidable). Moreover, we get from Proposition 2.12 that  $L \in Pol(C) \cap coPol(C)$  if and only if  $\alpha$  is a  $(Pol(C) \cap coPol(C))$ -morphism since  $Pol(C) \cap coPol(C)$  is a prevariety by Corollary 4.7. This property can be decided using (7) in Theorem 6.4 since we have  $\leq_C$  in hand.

**COROLLARY 6.5.** Let  $C$  be a positive prevariety with decidable membership. Then, membership is decidable for  $Pol(C) \cap coPol(C)$  as well.

This is surprising:  $C$ -membership suffices to handle  $(Pol(C) \cap coPol(C))$ -membership while we are only able to handle  $Pol(C)$ -membership when  $C$ -separation is decidable. Intuitively, this means that when a regular language  $L$  belongs to  $Pol(C) \cap coPol(C)$ , the basic languages in  $C$  which are required to construct  $L$  are all recognized by its syntactic morphism. On the other hand, this is not the case for  $Pol(C)$ .

We turn to the classes  $Pol(BPol(C)) \cap coPol(BPol(C))$ . As seen in Theorem 4.9, they are interesting because they correspond to the levels  $\Delta_2(\mathbb{I}_C)$  in quantifier alternation hierarchies.

We present a specialized characterization of them, which we prove as a corollary of Theorem 6.1 and Theorem 6.4. We start with a preliminary result. A key point is that we do not consider  $BPol$  in the proof: we bypass it using Lemma 4.8, which implies that  $Pol(BPol(C)) \cap coPol(BPol(C))$  and  $Pol(coPol(C)) \cap coPol(coPol(C))$  are the same class. Consequently, in order to apply Theorem 6.4, we need a description of the canonical preorder  $\leq_{coPol(C)}$  associated to a morphism. We use Theorem 6.1 to obtain one.

**LEMMA 6.6.** Let  $C$  be a positive prevariety and let  $\alpha : A^* \rightarrow M$  be a surjective morphism into a finite monoid. The relation  $\leq_{coPol(C)}$  on  $M$  is the least preorder such that for every  $x, y, s \in M$  and  $e \in E(M)$ , if  $(e, s) \in M^2$  is a  $C$ -pair, then  $xesy \leq_{coPol(C)} xey$ .

**PROOF.** One can check from the definition that  $\leq_{coPol(C)}$  is the reverse of the canonical preorder  $\leq_{Pol(C)}$  associated to  $\alpha$ . Therefore, it suffices prove that  $\leq_{Pol(C)}$  is the least preorder such that for every  $x, y, s \in M$  and  $e \in E(M)$ , if  $(e, s) \in M^2$  is a  $C$ -pair, then  $xey \leq_{Pol(C)} xesy$ .

By hypothesis on  $C$ , Theorem 4.3 implies that  $Pol(C)$  is a positive prevariety of regular languages. We first prove that for all  $x, y, s \in M$  and  $e \in E(M)$ , if  $(e, s) \in M^2$  is a  $C$ -pair, then  $xey \leq_{Pol(C)} xesy$ . Let  $f = [e]_{Pol(C)}$  and  $t = [s]_{Pol(C)}$ . Since  $\alpha$  is surjective, Lemma 5.19 yields that the map  $[\cdot]_{Pol(C)} \circ \alpha : A^* \rightarrow (M/\sim_{Pol(C)}, \leq_{Pol(C)})$  is a  $Pol(C)$ -morphism. Moreover, since  $e$  is an idempotent of  $M$ , it is immediate that  $f$  is an idempotent of  $M/\sim_{Pol(C)}$ . Finally, since  $(e, s)$  is a  $C$ -pair for  $\alpha$ , one can verify that  $(f, t) = ([e]_{Pol(C)}, [s]_{Pol(C)})$  is a  $C$ -pair for  $[\cdot]_{Pol(C)} \circ \alpha$ . Consequently, Theorem 6.1 yields  $f^{\omega+1} \leq_{Pol(C)} f^{\omega} t f^{\omega}$ , i.e.,  $f \leq_{Pol(C)} f t f$  since  $f$  is idempotent. This exactly says that  $[e]_{Pol(C)} \leq_{Pol(C)} [ese]_{Pol(C)}$ . Therefore, we also have  $[xey]_{Pol(C)} \leq_{Pol(C)} [xesy]_{Pol(C)}$  and we conclude that  $xey \leq_{Pol(C)} xesy$ , as desired.

Conversely, let  $\leq$  be the least preorder on  $M$  such that for every  $x, y, s \in M$  and  $e \in E(M)$ , if  $(e, s) \in M^2$  is a  $C$ -pair, then  $xey \leq xesy$ . We show that  $\leq_{Pol(C)} \subseteq \leq$ . Let  $\cong$  be equivalence on  $M$  induced by  $\leq$  and let  $\leq$  be the ordering induced by  $\leq$  on the quotient set  $M/\cong$ . It is straightforward that  $\leq$  is compatible with multiplication (this is where the elements  $x, y \in M$  in the definition are important). Hence, the pair  $(M/\cong, \leq)$  is an ordered monoid and the map  $s \mapsto [s]_{\cong}$ , sending every element  $s \in M$  to its  $\cong$ -class, is a morphism.

We use Theorem 6.1 to prove that  $[\cdot]_{\cong} \circ \alpha : A^* \rightarrow (M/\cong, \leq)$  is a  $Pol(C)$ -morphism. We have to show that given  $r, t \in M/\cong$  such that  $(r, t)$  is a  $C$ -pair for  $[\cdot]_{\cong} \circ \alpha$ , we have  $r^{\omega+1} \leq r^{\omega} t r^{\omega}$ . Let  $(r, t)$  be a  $C$ -pair for  $[\cdot]_{\cong} \circ \alpha$ . One can verify that there exists a  $C$ -pair  $(q, s) \in M^2$  for  $\alpha$  such that  $[q]_{\cong} = r$  and  $[s]_{\cong} = t$ . Let  $k = \omega(M)$ . We know from Lemma 5.12 that  $(q^k, s(q)^{k-1})$  is also a  $C$ -pair for  $\alpha$ . Since  $q^k \in E(M)$  by definition of  $k$ , we have  $q^k \leq q^k s q^{k-1} q^k$  by definition of  $\leq$  which yields  $q^{\omega+1} \leq q^{\omega} s q^{\omega}$  since  $k = \omega(M)$ . It follows that  $[q^{\omega+1}]_{\cong} \leq [q^{\omega} s q^{\omega}]_{\cong}$ . Therefore, since  $[q]_{\cong} = r$  and  $[s]_{\cong} = t$ , we obtain  $r^{\omega+1} \leq r^{\omega} t r^{\omega}$ , as desired, and we conclude that  $[\cdot]_{\cong} \circ \alpha : A^* \rightarrow (M/\cong, \leq)$  is a  $Pol(C)$ -morphism.

We are ready to prove that  $\leq_{Pol(C)} \subseteq \leq$ . Let  $(q, r) \in M$  such that  $q \leq_{Pol(C)} r$ . We show that  $q \leq r$ . Let  $F = \{p \in M \mid q \leq p\}$ . By definition,  $\alpha^{-1}(F)$  is recognized by the  $Pol(C)$ -morphism

$[\cdot]_{\cong} \circ \alpha : A^* \rightarrow (M/\cong, \leq)$ . Hence, we have  $\alpha^{-1}(F) \in \text{Pol}(C)$  and since we have  $q \in \alpha^{-1}(F)$  and  $q \leq_{\text{Pol}(C)} r$ , we obtain  $r \in F$  by definition of  $\leq_{\text{Pol}(C)}$ . This exactly says that  $q \leq r$  by definition of  $F$ , which concludes the proof.  $\blacksquare$

We are ready to present and prove the generic algebraic characterization of the classes  $\text{Pol}(\text{BPol}(C)) \cap \text{coPol}(\text{BPol}(C))$ .

**THEOREM 6.7.** *Let  $C$  be a positive prevariety and let  $\alpha : A^* \rightarrow M$  be a surjective morphism into a finite monoid. The two following properties are equivalent:*

1.  $\alpha$  is a  $(\text{Pol}(\text{BPol}(C)) \cap \text{coPol}(\text{BPol}(C)))$ -morphism.
2.  $\alpha$  satisfies the following property:

$$(eset)^{\omega+1} = (eset)^{\omega} et(eset)^{\omega} \quad (8)$$

for every  $s, t \in M$  and  $e \in E(M)$  such that  $(e, s) \in M^2$  is a  $C$ -pair.

**PROOF.** First, recall that by Lemma 4.8, we have the equality  $\text{Pol}(\text{BPol}(C)) = \text{Pol}(\text{coPol}(C))$ . Thus, the first assertion in the theorem states that  $\alpha$  is a  $(\text{Pol}(\text{coPol}(C)) \cap \text{coPol}(\text{coPol}(C)))$ -morphism. As  $\text{coPol}(C)$  is a positive prevariety by Corollary 4.7, it follows from Theorem 6.4 that this property holds if and only if  $\alpha$  satisfies the following condition:

$$q^{\omega+1} = q^{\omega} r q^{\omega} \quad \text{for every } q, r \in M \text{ such that } q \leq_{\text{coPol}(C)} r. \quad (9)$$

Consequently, it suffices to prove that  $\alpha$  satisfies (9) if and only if it satisfies (8).

We first assume that  $\alpha$  satisfies (9). We prove that it satisfies (8) as well. Given  $s, t \in M$  and  $e \in E(M)$  such that  $(e, s) \in M^2$  is a  $C$ -pair, we have to prove that  $(eset)^{\omega+1} = (eset)^{\omega} et(eset)^{\omega}$ . It is immediate from Lemma 6.6 that  $ese \leq_{\text{coPol}(C)} e$ , which yields  $eset \leq_{\text{coPol}(C)} et$ . Hence, (9) implies that  $(eset)^{\omega+1} = (eset)^{\omega} et(eset)^{\omega}$ , as desired.

Conversely, we assume that  $\alpha$  satisfies (8) and prove that it satisfies (9) as well. Given  $q, r \in M$  such that  $q \leq_{\text{coPol}(C)} r$ , we have to prove that  $q^{\omega+1} = q^{\omega} r q^{\omega}$ . By Lemma 6.6, there exists  $q_0, \dots, q_n \in M$  such that  $q = q_0$ ,  $r = q_n$  and, for every  $i \leq n$ , there exist  $x, y, s \in M$  and  $e \in E(M)$  such that  $(e, s) \in M^2$  is a  $C$ -pair,  $q_{i-1} = xesey$  and  $q_i = xey$ . We use induction on  $i$  to prove that  $q^{\omega+1} = q^{\omega} q_i q^{\omega}$  for every  $i \leq n$ . Since  $q_n = r$ , the case  $i = n$  yields the desired result. When  $i = 0$ , it is immediate that  $q^{\omega+1} = q^{\omega} q_0 q^{\omega}$  since  $q_0 = q$ . Assume now that  $i \geq 1$ . By induction hypothesis, we have  $q^{\omega+1} = q^{\omega} q_{i-1} q^{\omega}$ , which implies that  $q^{\omega+2} = (q^{\omega} q_{i-1} q^{\omega})^{\omega+2}$ . Moreover, by definition, we have  $x, y, s \in M$  and  $e \in E(M)$  such that  $(e, s) \in M^2$  is a  $C$ -pair,  $q_{i-1} = xesey$  and  $q_i = xey$ . It follows that,

$$\begin{aligned} q^{\omega+2} &= (q^{\omega} xesey q^{\omega})^{\omega+2} \\ &= q^{\omega} x(esey q^{\omega} q^{\omega} x)^{\omega+1} esey q^{\omega}. \end{aligned}$$

Since  $(e, s) \in M^2$  is a  $C$ -pair, it now follows from (8) applied with  $t = yq^\omega q^\omega x$  that,

$$\begin{aligned} q^{\omega+2} &= q^\omega x (ese y q^\omega q^\omega x)^\omega e y q^\omega q^\omega x (ese y q^\omega q^\omega x)^\omega ese y q^\omega \\ &= (q^\omega x ese y q^\omega)^\omega q^\omega x e y q^\omega (q^\omega x ese y q^\omega)^{\omega+1} \\ &= (q^\omega q_{i-1} q^\omega)^\omega q^\omega q_i q^\omega (q^\omega q_{i-1} q^\omega)^{\omega+1} \\ &= q^\omega q_i q^{\omega+1}. \end{aligned}$$

Multiplying by  $q^{\omega-1}$  on the right yields  $q^{\omega+1} = q^\omega q_i q^\omega$ , which concludes the proof.  $\blacksquare$

Theorem 6.7 implies that membership is decidable for  $Pol(BPol(C)) \cap coPol(BPol(C))$  when  $C$  is a positive prevariety with decidable separation. The argument is based on Proposition 2.12. Note that we already had this result. Indeed, since  $Pol(BPol(C)) \cap coPol(BPol(C))$  is exactly  $Pol(coPol(C)) \cap coPol(coPol(C))$  by Lemma 4.8, it also follows from Corollary 6.2 and Corollary 6.5 (clearly, membership for  $coPol(C)$  reduces to the same problem for  $Pol(C)$ ). Yet, we shall need Theorem 6.7 when proving the logical characterizations of  $UPol$  in Section 10.

**COROLLARY 6.8.** Let  $C$  be a positive prevariety with decidable separation. Then, membership is decidable for  $Pol(BPol(C)) \cap coPol(BPol(C))$ .

## 7. Algebraic characterizations for unambiguous polynomial closure

We present the generic algebraic characterization of *unambiguous* polynomial closure. It holds for every input *prevariety*  $C$  and implies that  $UPol(C)$ -membership reduces to the same problem for  $C$ . We also use this characterization to complete the proofs of Theorem 4.21 and Theorem 4.29: when  $C$  is a prevariety, we have  $UPol(C) = APol(C) = Pol(C) \cap coPol(C)$ .

**REMARK 7.1.** There exists an independent characterization by Pin [15]. Yet, it involves algebraic notions that we have not presented and does not yield the reduction to  $C$ -membership.

Again, the characterization is based on the canonical equivalence  $\sim_C$  associated to morphisms into finite monoids (see Section 5 for the definition).

**THEOREM 7.2.** Let  $C$  be a prevariety and let  $\alpha : A^* \rightarrow M$  be a surjective morphism into a finite monoid. The following properties are equivalent:

1.  $\alpha$  is a  $UPol(C)$ -morphism.
2.  $\alpha$  is a  $APol(C)$ -morphism.
3.  $\alpha$  is a  $WAPol(C)$ -morphism.
4.  $\alpha$  satisfies the following property:

$$s^{\omega+1} = s^\omega t s^\omega \quad \text{for every } C\text{-pair } (s, t) \in M^2. \quad (6)$$

5.  $\alpha$  satisfies the following property:

$$s^{\omega+1} = s^\omega t s^\omega \quad \text{for every } s, t \in M \text{ such that } s \sim_C t. \quad (10)$$

Note that by Theorem 6.4, Equation (6) characterizes  $(Pol(C) \cap coPol(C))$ -morphisms. Before proving Theorem 7.2, let us discuss its consequences. Importantly, we cannot directly apply it to reduce  $UPol(C)$ -membership to  $C$ -membership. Indeed, given an arbitrary class  $\mathcal{D}$ , the connection between  $\mathcal{D}$ -membership and  $\mathcal{D}$ -morphisms is based on Proposition 2.12, which requires  $\mathcal{D}$  to be a (positive) prevariety. Hence, we need Theorem 4.24:  $UPol(C)$  is a prevariety when  $C$  is one. Yet, we cannot use Theorem 4.24 at this point, since we obtained it as the corollary of a result that we have not proved yet: Theorem 4.21 (which states that  $UPol(C) = Pol(C) \cap coPol(C)$  when  $C$  is a prevariety). Therefore, we first use Theorem 7.2 to provide a proof of Theorem 4.21. In fact, we present a more general statement involving also  $WAPol(C)$  and  $APol(C)$ . It combines Theorem 4.21 and Theorem 4.29.

**COROLLARY 7.3.** For any prevariety  $C$ , we have the following equalities:

$$WAPol(C) = APol(C) = UPol(C) = Pol(C) \cap coPol(C).$$

**PROOF.** The inclusions  $WAPol(C) \subseteq APol(C) \subseteq UPol(C)$  are immediate from Lemma 4.14: left/right deterministic marked concatenations are unambiguous. Hence, we need to prove that  $UPol(C) \subseteq WAPol(C)$ ,  $UPol(C) \subseteq Pol(C) \cap coPol(C)$  and  $Pol(C) \cap coPol(C) \subseteq UPol(C)$ . We start with the latter. Consider a language  $L \in Pol(C) \cap coPol(C)$  and let  $\alpha : A^* \rightarrow M$  be the syntactic morphism of  $L$ . Since  $Pol(C) \cap coPol(C)$  is a prevariety by Corollary 4.7, it follows from Proposition 2.12 that  $\alpha$  is a  $(Pol(C) \cap coPol(C))$ -morphism. It then follows from Theorem 6.4 that  $\alpha$  satisfies (6). Consequently, Theorem 7.2 yields that  $\alpha$  is a  $UPol(C)$ -morphism, and  $L \in UPol(C)$  since it is recognized by its syntactic morphism.

We turn to the inclusions  $UPol(C) \subseteq WAPol(C)$  and  $UPol(C) \subseteq Pol(C) \cap coPol(C)$ , which we treat simultaneously. Let  $L \in UPol(C)$  and  $\alpha : A^* \rightarrow M$  be the syntactic morphism of  $L$ . We prove that  $\alpha$  satisfies (6). It will follow from Theorem 6.4 that  $\alpha$  is a  $Pol(C) \cap coPol(C)$ -morphism and from Theorem 7.2 that it is a  $WAPol(C)$ -morphism. Therefore, since  $L$  is recognized by its syntactic morphism, we get  $L \in Pol(C) \cap coPol(C)$  and  $L \in WAPol(C)$ .

In order to prove (6), let  $(s, t) \in M^2$  be a  $C$ -pair for  $\alpha$ . We have to show that  $s^{\omega+1} = s^\omega t s^\omega$ . Since  $L \in UPol(C)$ , Proposition 4.26 yields a  $C$ -morphism  $\eta : A^* \rightarrow N$  and  $k \in \mathbb{N}$  such that for all words  $u, v, v', x, y \in A^*$  satisfying  $\eta(u) = \eta(v) = \eta(v')$ , we have  $xu^k v u^k y \in L \Leftrightarrow xu^k v' u^k y \in L$ . Since  $(s, t)$  is a  $C$ -pair and  $\eta$  is a  $C$ -morphism, it follows from Lemma 5.11 that there exist  $u, v \in A^*$  such that  $\eta(u) = \eta(v)$ ,  $\alpha(u) = s$  and  $\alpha(v) = t$ . By definition of  $\eta$ , it follows that for every  $x, y \in A^*$ , we have  $xu^k v u^k y \in L \Leftrightarrow xu^k u u^k y \in L$ . This exactly says that  $u^k v u^k$  and  $u^k u u^k$  are equivalent for the syntactic congruence of  $L$ . Hence, they have the same image under  $\alpha$  by definition of the syntactic morphism, and we get  $s^k t s^k = s^k s s^k$ . It now suffices to multiply by the appropriate number of copies of  $s$  on the left and on the right to get  $s^{\omega+1} = s^\omega t s^\omega$ , which completes the proof. ■

**REMARK 7.4.** There is a subtle difference between Theorem 7.2 above and Theorem 6.4 (which applies to  $Pol(C) \cap coPol(C)$ ), although the algebraic characterizations are the same: the latter holds for *every positive prevariety*  $C$ , while the former requires  $C$  to be a *prevariety*. This is important, as Corollary 7.3 fails when  $C$  is only a positive prevariety (see Remark 4.22).

With Corollary 7.3 in hand, we may now use Theorem 4.24: if  $C$  is a prevariety, then so is  $UPol(C)$ . Consequently, in this case, it follows from Proposition 2.12 that deciding whether an input regular language  $L$  belongs to  $UPol(C)$  amounts to testing whether its syntactic morphism is a  $UPol(C)$ -morphism. The syntactic morphism of a regular language can be computed. Moreover, Theorem 7.2 states that an arbitrary surjective morphism is a  $UPol(C)$ -morphism if and only if it satisfies (10). This can be checked by testing every possible combination provided that we have the equivalence  $\sim_C$  in hand. Finally, by Lemma 5.14,  $\sim_C$  can be computed as soon as  $C$ -membership is decidable. Altogether, we obtain that  $UPol$  preserves the decidability of membership when it is applied to a prevariety.

**COROLLARY 7.5.** Let  $C$  be a prevariety. Assume that  $C$  has decidable membership. Then,  $UPol(C)$ -membership is decidable as well.

It remains to establish Theorem 7.2. We devote the rest of this section to this proof.

**PROOF OF THEOREM 7.2.** We fix a prevariety  $C$  and a surjective morphism  $\alpha : A^* \rightarrow M$ . We prove (3)  $\Rightarrow$  (2)  $\Rightarrow$  (1)  $\Rightarrow$  (4)  $\Rightarrow$  (5)  $\Rightarrow$  (3). The implications (3)  $\Rightarrow$  (2) and (2)  $\Rightarrow$  (1) are clear: we have the inclusions  $WAPol(C) \subseteq APol(C) \subseteq UPol(C)$  since left/right deterministic marked concatenations are necessarily unambiguous by Lemma 4.14.

We show (1)  $\Rightarrow$  (4): let  $\alpha$  be a  $UPol(C)$ -morphism, we prove that it satisfies (6). Given a  $C$ -pair  $(s, t) \in M^2$ , we show that  $s^{\omega+1} = s^\omega t s^\omega$ . Let  $q = s^\omega$ . By hypothesis, we have  $\alpha^{-1}(q) \in UPol(C)$ . Hence, Proposition 4.26 yields a  $C$ -morphism  $\eta : A^* \rightarrow N$  and  $k \in \mathbb{N}$  such that for every  $u, v, v' \in A^*$ , if  $\eta(u) = \eta(v) = \eta(v')$ , then  $u^k v u^k \in \alpha^{-1}(q) \Leftrightarrow u^k v' u^k \in \alpha^{-1}(q)$ . Let  $p = \omega(M)$ . Since  $(s, t) \in M^2$  is a  $C$ -pair, it follows from Lemma 5.12 that  $(s^p, t s^{p-1})$  is also a  $C$ -pair. Hence, since  $\eta$  is a  $C$ -morphism, Lemma 5.11 yields  $u, v \in A^*$  such that  $\eta(u) = \eta(v)$ ,  $\alpha(u) = s^p$  and  $\alpha(v) = t s^{p-1}$ . Clearly,  $\alpha(u^k u u^k) = q^{kp} q^p q^{kp} = q = s^\omega$  since  $p = \omega(M)$ . Hence  $u^k u u^k \in \alpha^{-1}(q)$  and since  $\eta(u) = \eta(v)$ , the definition of  $\eta$  yields  $u^k v u^k \in \alpha^{-1}(q)$ . We get  $\alpha(u^k v u^k) = q = s^\omega$ . By definition of  $u$  and  $v$ , this exactly says that  $s^\omega t s^{2\omega-1} = s^\omega$ . It now suffices to multiply by  $s$  on the right to get  $s^{\omega+1} = s^\omega t s^\omega$ , as desired.

We now prove (4)  $\Rightarrow$  (5). Assume that  $\alpha$  is surjective and satisfies (6), we need to prove that it satisfies (10) as well. Since  $C$  is a prevariety, we know that  $\sim_C = \leq_C$  by Lemma 5.17. Hence, the property is immediate from the implication (2)  $\Rightarrow$  (3) in Theorem 6.4.

We proved that (3)  $\Rightarrow$  (2)  $\Rightarrow$  (1)  $\Rightarrow$  (4)  $\Rightarrow$  (5). Therefore, it remains to prove (5)  $\Rightarrow$  (3). Assuming that  $\alpha$  satisfies (10), we show that it is a  $WAPol(C)$ -morphism, *i.e.*, that every language

recognized by  $\alpha$  belongs to  $WAPol(C)$ . Let  $N = M/\sim_C$ . Recall that  $N$  is a monoid since  $\sim_C$  is a congruence by Lemma 5.18. We define  $\eta$  as the map  $\eta = [\cdot]_C \circ \alpha : A^* \rightarrow N$ . We know from Lemma 5.19 that  $\eta$  is a  $C$ -morphism. Given a finite set of languages  $\mathbf{K}$  and  $s, t \in M$ , we say that  $\mathbf{K}$  is  $(s, t)$ -safe if for all  $K \in \mathbf{K}$  and  $w, w' \in K$ , we have  $s\alpha(w)t = s\alpha(w')t$ . Finally, a  $WAPol(C)$ -partition of a language  $H$  is a finite partition of  $H$  into languages of  $WAPol(C)$ . The implication (5)  $\Rightarrow$  (3) follows from the following lemma.

**LEMMA 7.6.** For any  $x \in N$  and  $s, t \in M$ , there exists an  $(s, t)$ -safe  $WAPol(C)$ -partition of  $\eta^{-1}(x)$ .

We first apply Lemma 7.6 to complete the main argument. For every  $x \in N$ , Lemma 7.6 yields a  $WAPol(C)$ -partition  $\mathbf{K}_x$  of  $\eta^{-1}(x)$  which is  $(1_M, 1_M)$ -safe. Hence,  $\mathbf{K} = \bigcup_{x \in N} \mathbf{K}_x$  is a  $WAPol(C)$ -partition of  $A^*$  which is  $(1_M, 1_M)$ -safe. The latter property implies that for every  $K \in \mathbf{K}$ , we have  $s \in M$  such that  $K \subseteq \alpha^{-1}(s)$ . Since  $\mathbf{K}$  is a partition of  $A^*$ , it follows that every language recognized by  $\alpha$  is a disjoint finite union of languages in  $\mathbf{K}$  and therefore belongs to  $WAPol(C)$ , by closure under disjoint union. This concludes the main argument.

It remains to prove Lemma 7.6. Let  $s, t \in M$  and  $x \in N$ . We build an  $(s, t)$ -safe  $WAPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(x)$ . The proof proceeds by induction on the following three parameters, which depend on Green relations on  $M$  and  $N$ , listed by order of importance:

1. The *rank of*  $(s, x, t)$ : the number of elements  $y \in N$  such that  $[s]_C x [t]_C \leq_{\mathcal{J}} y$ .
2. The  *$\mathcal{R}$ -index of*  $s$ : the number of elements  $r \in M$  such that  $r \leq_{\mathcal{R}} s$
3. The  *$\mathcal{L}$ -index of*  $t$ : the number of elements  $r \in M$  such that  $r \leq_{\mathcal{L}} t$ .

There are three cases, depending on whether  $[s]_C x [t]_C \mathcal{J} x$  and on two properties of  $s, t$  and  $x$ :

- We say that  $s$  is *right  $x$ -stable* if there exists  $q \in M$  such that  $[q]_C \mathcal{R} x$  and  $sq \mathcal{R} s$ .
- We say that  $t$  is *left  $x$ -stable* if there exists  $r \in M$  such that  $[r]_C \mathcal{L} x$  and  $rt \mathcal{L} t$ .

In the base case, we assume that all three properties hold. Otherwise, we distinguish two inductive cases depending on whether  $[s]_C x [t]_C \mathcal{J} x$  holds or not.

**Base case:**  $[s]_C x [t]_C \mathcal{J} x$ ,  $s$  is right  $x$ -stable and  $t$  is left  $x$ -stable. In this case, let  $\mathbf{K} = \{\eta^{-1}(x)\}$ , which is clearly a  $WAPol(C)$ -partition of  $\eta^{-1}(x)$ : this is even a  $C$ -partition of  $\eta^{-1}(x)$  since  $\eta$  is a  $C$ -morphism. It remains to show that  $\mathbf{K}$  is  $(s, t)$ -safe. Given  $w, w' \in \eta^{-1}(x)$ , show that  $s\alpha(w)t = s\alpha(w')t$ . We let  $p = \alpha(w)$  and  $p' = \alpha(w')$ . By hypothesis, we have  $[p]_C = [p']_C = x$  and we have to show that  $spt = sp't$ . We have  $[s]_C x [t]_C \mathcal{J} x$ , which implies that  $[s]_C x \mathcal{J} x$  and  $x[t]_C \mathcal{J} x$ . Hence, since  $[s]_C x \leq_{\mathcal{L}} x$  and  $x[t]_C \leq_{\mathcal{R}} x$ , Lemma 2.1 yields  $[s]_C x \mathcal{L} x$  and  $x[t]_C \mathcal{R} x$ . We use this property to prove the following fact.

**LEMMA 7.7.** We have  $spt \mathcal{R} s$  and  $spt \mathcal{L} t$ .

**PROOF.** We prove that  $spt \mathcal{R} s$  using the hypotheses that  $x[t]_C \mathcal{R} x$  and that  $s$  is right  $x$ -stable (that  $spt \mathcal{L} t$  is proved symmetrically using the hypotheses that  $[s]_C x \mathcal{L} x$  and that  $t$  is left

$x$ -stable). Since  $s$  is right  $x$ -stable, there exists  $q \in M$  such that  $[q]_C \mathcal{R} x$  and  $sq \mathcal{R} s$ . Since  $x = [p]_C$  and  $x[t]_C \mathcal{R} x$ , we get  $[pt]_C \mathcal{R} [q]_C$ . We get  $r \in M$  such that  $[q]_C = [ptr]_C$ . Moreover, since  $sq \mathcal{R} s$ , we have  $q' \in M$  such that  $s = sqq'$ . Consequently, we get  $s = s(qq')^\omega = s(qq')^{\omega+1}$ . We have  $[qq']_C = [ptrq']_C$  which implies that  $qq' \sim_C ptrq'$ . Therefore, Equation (10) yields  $(qq')^{\omega+1} = (qq')^\omega ptrq'(qq')^\omega$ . Finally, we may multiply by  $s$  on the left to obtain,

$$s = s(qq')^{\omega+1} = s(qq')^\omega ptrq'(qq')^\omega = sptrq'(qq')^\omega.$$

This implies that  $s \leq_{\mathcal{R}} spt$ . Since it is clear that  $spt \leq_{\mathcal{R}} s$ , we get  $spt \mathcal{R} s$ , as desired.  $\blacksquare$

Lemma 7.7 yields  $q_1, q_2 \in M$  such that  $t = q_1spt$  and  $s = sptq_2$ . Let  $r = q_1sptq_2$ . We may combine the above equalities to obtain  $s = spr$  (indeed,  $r = tq_2$ , whence  $spr = sptq_2 = s$ ). Therefore,  $s = s(pr)^\omega$ . Similarly,  $t = rpt = (rp)^{\omega+1}t$ . By hypothesis,  $[p]_C = [p']_C = x$ . Hence, we obtain  $[pr]_C = [p'r]_C$ , so that  $pr \sim_C p'r$ . Therefore, Equation (10) applies and yields,

$$(pr)^{2\omega+1} = (pr)^{\omega+1} = (pr)^\omega p'r(pr)^\omega.$$

We may now multiply by  $s$  on the left and by  $pt$  on the right to obtain,

$$s(pr)^\omega p(rp)^{\omega+1}t = s(pr)^\omega p'(rp)^{\omega+1}t.$$

Since we already established that  $s = s(pr)^\omega$  and  $t = (rp)^{\omega+1}t$ , we get as desired that  $spr = sp't$ , which concludes the proof for this base case.

**First inductive case:**  $[s]_C x [t]_C <_{\mathcal{J}} x$ . In this case, it is immediate that the rank of  $(1_M, x, 1_M)$  is strictly smaller than the one of  $(s, x, t)$ . Hence, induction on our first and main parameter in Lemma 7.6 yields a  $WAPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(x)$  which is  $(1_M, 1_M)$ -safe (and therefore  $(s, t)$ -safe as well). This concludes the first inductive case.

**Second inductive case:**  $[s]_C x [t]_C \mathcal{J} x$  and either  $s$  is not right  $x$ -stable or  $t$  is not left  $x$ -stable. This involves two symmetrical arguments depending on which property holds. We treat the case where  $t$  is not left  $x$ -stable using induction on the first and third parameter (the second one is used in the symmetrical case). Let  $T \subseteq N \times A \times N$  be the set of all triples  $(y, a, z) \in N \times A \times N$  such that  $x \mathcal{L} \eta(a)z <_{\mathcal{L}} z$  and  $y\eta(a)z = x$ . For each triple  $(y, a, z) \in T$ , we use induction to build auxiliary  $WAPol(C)$ -partitions of  $\eta^{-1}(y)$  and  $\eta^{-1}(z)$ , which we then combine to construct the desired  $(s, t)$ -safe  $WAPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(x)$ . We fix  $(y, a, z) \in T$  for the definition of these auxiliary  $WAPol(C)$ -partitions.

By definition of  $T$ , we have  $x <_{\mathcal{L}} z$ . By Lemma 2.1, this implies  $x <_{\mathcal{J}} z$ , so that we also have  $[s]_C x [t]_C <_{\mathcal{J}} z$ . Hence, the rank of  $(1_M, z, 1_M)$  is strictly smaller than the one of  $(s, x, t)$ . Since this is our most important induction parameter, we obtain a  $WAPol(C)$ -partition  $\mathbf{V}_z$  of  $\eta^{-1}(z)$  which is  $(1_M, 1_M)$ -safe.

We now use  $\mathbf{V}_z$  to build several  $WAPol(C)$ -partitions of  $\eta^{-1}(y)$ , one for each language  $V \in \mathbf{V}_z$ . We fix a language  $V \in \mathbf{V}_z$  for the definition. Since  $\mathbf{V}$  is  $(1_M, 1_M)$ -safe, there exists  $r_V \in M$  such that  $\alpha(v) = r_V$  for every  $v \in V$ . Moreover,  $[r_V]_C = \eta(v) = z$  since  $v \in \eta^{-1}(z)$ . We show that  $(s, y, \alpha(a)r_V t)$  has a strictly smaller induction parameter than  $(s, x, t)$ . Since  $(y, a, z) \in T$ , we have  $x \mathcal{L} \eta(a)z$ , which means that  $x \mathcal{L} [\alpha(a)r_V]_C$ . Since  $t$  is not left  $x$ -stable, it follows that  $\alpha(a)r_V t <_{\mathcal{L}} t$ : the  $\mathcal{L}$ -index of  $\alpha(a)r_V t$  is strictly smaller than the one of  $t$ , meaning that our third induction parameter has decreased. Moreover, since  $y\eta(a)z = x$ , we also have  $[s]_C y [\alpha(a)r_V t]_C = [s]_C x [t]_C$  which means that  $(s, x, t)$  and  $(s, y, \alpha(a)r_V t)$  have the same rank: our first induction parameter is unchanged. Finally, the second induction parameter remains unchanged as well, since it only depends on  $s$ . Hence, induction on our third parameter yields a  $WAPol(C)$ -partition  $\mathbf{U}_{(y,a,z),V}$  of  $\eta^{-1}(y)$  which is  $(s, \alpha(a)r_V t)$ -safe.

We are ready to build the desired  $(s, t)$ -safe  $WAPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(x)$ . We define,

$$\mathbf{K} = \bigcup_{(y,a,z) \in T} \{UaV \mid V \in \mathbf{V}_z \text{ and } U \in \mathbf{U}_{(y,a,z),V}\}.$$

It remains to prove that  $\mathbf{K}$  satisfies the desired properties. First, we show that  $\mathbf{K}$  is indeed a  $WAPol(C)$ -partition of  $\eta^{-1}(x)$ . We start by proving that for every  $w \in \eta^{-1}(x)$ , there exists a *unique*  $K \in \mathbf{K}$  such that  $w \in K$ . Let  $v'$  be the least suffix of  $w$  such that  $\eta(v') \mathcal{L} x$  ( $u'$  exists since  $w$  is such a suffix). Observe that  $v' \neq \varepsilon$ . Indeed, otherwise  $[1_M]_C \mathcal{L} x$  and  $1_M t = t \mathcal{L} t$ , contradicting our hypothesis that  $t$  is not left  $x$ -stable. Thus,  $v' = av$  for some  $v \in A^*$  and  $a \in A$ . Finally, we let  $u \in A^*$  be such that  $w = uav$ . Let then  $y = \eta(u)$  and  $z = \eta(v)$ . By definition of  $v' = av$ , we know that  $x \mathcal{L} \eta(a)z <_{\mathcal{L}} z$  and  $y\eta(a)z = x$ . Hence,  $(y, a, z) \in T$ . Moreover, there exists  $V \in \mathbf{V}_z$  such that  $v \in V$  and  $U \in \mathbf{U}_{(y,a,z),V}$  such that  $u \in U$ . Thus  $w = uav \in UaV$  and  $UaV \in \mathbf{K}$ . Moreover, one can verify from the definition that this is the only language in  $\mathbf{K}$  containing  $w$ .

We now prove that for every  $K \in \mathbf{K}$ , we have  $K \subseteq \eta^{-1}(x)$  and  $K \in WAPol(C)$ . By definition,  $K = UaV$  with  $V \in \mathbf{V}_y$  and  $U \in \mathbf{U}_{(y,a,z),V}$  for some triple  $(y, a, z) \in T$ . Moreover, by definition of  $\mathbf{V}_y$  and  $\mathbf{U}_{(y,a,z),V}$ , we have  $U \subseteq \eta^{-1}(y)$  and  $V \subseteq \eta^{-1}(z)$ . Thus, since we have  $y\eta(a)z = x$  by definition of  $T$ , we get  $K = UaV \subseteq \eta^{-1}(x)$ . Moreover,  $U, V \in WAPol(C)$ . Therefore, it suffices to show that  $UaV$  is right  $C$ -deterministic to prove that  $UaV \in WAPol(C)$  (note that left  $C$ -deterministic marked concatenations are used in the symmetrical case). By definition, we have  $V \subseteq \eta^{-1}(y)$  and  $U \subseteq \eta^{-1}(z)$ . Hence, since  $\eta^{-1}(y), \eta^{-1}(z) \in C$  since  $\eta$  is a  $C$ -morphism, it suffices to prove that  $\eta^{-1}(y)a\eta^{-1}(z)$  is right deterministic. This is immediate from Lemma 4.15: since  $(y, a, z) \in T$  we have  $\eta(a)z <_{\mathcal{L}} z$ . Altogether, we conclude that  $\mathbf{K}$  is indeed a  $WAPol(C)$ -partition of  $\eta^{-1}(x)$ .

Finally, we show that  $\mathbf{K}$  is  $(s, t)$ -safe. Let  $K \in \mathbf{K}$  and  $w, w' \in K$ . We have to prove that  $s\alpha(w)t = s\alpha(w')t$ . By definition  $K = UaV$  with  $V \in \mathbf{V}_z$  and  $U \in \mathbf{U}_{(y,a,z),V}$  for some triple  $(y, a, z) \in T$ . Thus, we get  $u, u' \in U$  and  $v, v' \in V$  such that  $w = uav$  and  $w' = u'av'$ . By

definition  $v, v' \in \mathbf{V}$  imply that  $\alpha(v) = \alpha(v') = r_V$ . Therefore, we obtain  $s\alpha(w)t = s\alpha(u)\alpha(a)r_Vt$  and  $s\alpha(w')t = s\alpha(u')\alpha(a)r_Vt$ . Finally, recall that  $\mathbf{U}_V$  is  $(s, \alpha(a)r_Vt)$ -safe. Consequently, we get  $s\alpha(w)t = s\alpha(u)\alpha(a)r_Vt = s\alpha(u')\alpha(a)r_Vt = s\alpha(w')t$ , which concludes the proof. ■

## 8. Covering and Separation for unambiguous polynomial closure

In this section, we look at separation and covering for classes of the form  $UPol(C)$ . We prove that both problems are decidable for  $UPol(C)$  when the input class  $C$  is a *finite* prevariety. The algorithm is based on a generic framework introduced in [32] to handle separation and covering. We first recall this framework and then use it to present the algorithm.

### 8.1 Semirings

A *semiring* is a tuple  $(R, +, \cdot)$  where  $R$  is a set and “+” and “ $\cdot$ ” are two binary operations called addition and multiplication, which satisfy the following axioms:

- $(R, +)$  is a commutative monoid, whose identity element is denoted by  $0_R$ .
- $(R, \cdot)$  is a monoid, whose identity element is denoted by  $1_R$ .
- Multiplication distributes over addition: for  $r, s, t \in R$ ,  $r \cdot (s + t) = (r \cdot s) + (r \cdot t)$  and  $(r + s) \cdot t = (r \cdot t) + (s \cdot t)$ .
- $0_R$  is a zero for  $(R, \cdot)$ :  $0_R \cdot r = r \cdot 0_R = 0_R$  for every  $r \in R$ .

A semiring  $R$  is *idempotent* when  $r + r = r$  for every  $r \in R$ , *i.e.*, when the additive monoid  $(R, +)$  is idempotent (there is no additional constraint on the multiplicative monoid  $(R, \cdot)$ ). Given an idempotent semiring  $(R, +, \cdot)$ , one can define a canonical ordering  $\leq$  over  $R$ :

$$\text{For all } r, s \in R, \quad r \leq s \text{ when } r + s = s.$$

It is easy to check that  $\leq$  is a partial order which is compatible with both addition and multiplication. Moreover, every morphism between two such commutative and idempotent monoids is increasing for this ordering.

**EXAMPLE 8.1.** A key example of idempotent semiring is the set of all languages  $2^{A^*}$ . Union is the addition and language concatenation is the multiplication (with  $\{\varepsilon\}$  as the identity element). Observe that in this case, the canonical ordering is inclusion. More generally, if  $M$  is a monoid, then  $2^M$  is an idempotent semiring whose addition is union, and whose multiplication is obtained by pointwise lifting that of  $M$  to subsets.

When dealing with subsets of an idempotent semiring  $R$ , we shall often apply a *downset operation*. Given  $S \subseteq R$ , we write:

$$\downarrow_R S = \{r \in R \mid r \leq s \text{ for some } s \in S\}.$$

We extend this notation to Cartesian products of arbitrary sets with  $R$ . Given some set  $X$  and  $S \subseteq X \times R$ , we write,

$$\downarrow_R S = \{(x, r) \in X \times R \mid \exists s \in R \text{ such that } r \leq s \text{ and } (x, s) \in S\}.$$

**Multiplicative rating maps** We define a *multiplicative rating map* as a semiring morphism  $\rho : (2^{A^*}, \cup, \cdot) \rightarrow (R, +, \cdot)$  where  $(R, +, \cdot)$  is a *finite* idempotent semiring, called the *rating set of  $\rho$* . That is,  $\rho$  is a map from  $2^{A^*}$  to  $R$  satisfying the following properties:

1.  $\rho(\emptyset) = 0_R$ .
2. For every  $K_1, K_2 \subseteq A^*$ , we have  $\rho(K_1 \cup K_2) = \rho(K_1) + \rho(K_2)$ .
3.  $\rho(\{\varepsilon\}) = 1_R$ .
4. For every  $K_1, K_2 \subseteq A^*$ , we have  $\rho(K_1 K_2) = \rho(K_1) \cdot \rho(K_2)$ .

For the sake of improved readability, when applying a multiplicative rating map  $\rho$  to a singleton set  $\{w\}$ , we shall write  $\rho(w)$  for  $\rho(\{w\})$ . Additionally, we write  $\rho_* : A^* \rightarrow R$  for the restriction of  $\rho$  to  $A^*$ : for every  $w \in A^*$ , we have  $\rho_*(w) = \rho(w)$  (this notation is useful when referring to the language  $\rho_*^{-1}(r) \subseteq A^*$ , which consists of all words  $w \in A^*$  such that  $\rho(w) = r$ ). Note that  $\rho_*$  is a morphism into the finite monoid  $(R, \cdot)$ .

**REMARK 8.2.** As the adjective “multiplicative” suggests, there exists a more general notion of “rating map” introduced in [32]. These are morphisms of idempotent and commutative monoids (*i.e.*, it is not required that  $R$  be equipped with a multiplication). However, we shall not use this notion in the paper.

Most of the theory makes sense for arbitrary multiplicative rating maps. Yet, in the paper, we work with special multiplicative rating maps satisfying an additional property.

**Finitary and full rating maps.** We say that a multiplicative rating map  $\rho : 2^{A^*} \rightarrow R$  is *finitary* when, for every language  $K \subseteq A^*$ , there exist finitely many words  $w_1, \dots, w_n \in K$  such that  $\rho(K) = \rho(w_1) + \dots + \rho(w_n)$ .

If a rating map is *simultaneously* finitary and multiplicative, we say that it is *full*. Full rating maps are specially important. This is because any full rating map  $\rho : 2^{A^*} \rightarrow R$  is (fully) characterized by the canonical monoid morphism  $\rho_* : A^* \rightarrow R$ . Indeed, for  $K \subseteq A^*$ , we may consider the sum of all elements  $\rho(w)$  for  $w \in K$ : while it may be infinite, this sum boils down to a finite one since  $R$  is commutative and idempotent for addition. The hypothesis that  $\rho$  is finitary implies that  $\rho(K)$  is equal to this sum. The key point here is that full rating maps are finitely representable: clearly, a full rating map  $\rho$  is characterized by the morphism  $\rho_* : A^* \rightarrow R$ , which is finitely representable since it is a morphism into a finite monoid. Hence, we may speak about algorithms taking full rating maps as input.

**Canonical multiplicative rating map associated to a monoid morphism.** Finally, one can associate a particular full rating map  $\rho_\alpha$  to every monoid morphism  $\alpha : A^* \rightarrow M$  into a finite monoid. Its rating set is the idempotent semiring  $(2^M, \cup, \cdot)$ , whose multiplication is obtained by lifting the one of  $M$  to subsets of  $M$ . Moreover, for every language  $K \subseteq A^*$ , we let  $\rho_\alpha(K)$  be the direct image  $\alpha(K) \subseteq M$ . In other words, we define:

$$\begin{aligned} \rho_\alpha : 2^{A^*} &\rightarrow 2^M \\ K &\mapsto \{\alpha(w) \mid w \in K\}. \end{aligned}$$

Clearly,  $\rho_\alpha$  is a full rating map.

**Optimal imprints.** Now that we have multiplicative rating maps, we turn to imprints. Consider a multiplicative rating map  $\rho : 2^{A^*} \rightarrow R$ . Given any finite set of languages  $\mathbf{K}$ , we define the  $\rho$ -imprint of  $\mathbf{K}$ . Intuitively, when  $\mathbf{K}$  is a cover of some language  $L$ , this object measures the “quality” of  $\mathbf{K}$ . The  $\rho$ -imprint of  $\mathbf{K}$  is the subset of  $R$  defined by:

$$\mathcal{I}[\rho](\mathbf{K}) = \downarrow_R \{\rho(K) \mid K \in \mathbf{K}\}.$$

We now define optimality. Consider an arbitrary multiplicative rating map  $\rho : 2^{A^*} \rightarrow R$  and a lattice  $\mathcal{D}$ . Given a language  $L$ , an *optimal  $\mathcal{D}$ -cover* of  $L$  for  $\rho$  is a  $\mathcal{D}$ -cover  $\mathbf{K}$  of  $L$  having the smallest possible imprint among all  $\mathcal{D}$ -covers, i.e., which satisfies the following property:

$$\mathcal{I}[\rho](\mathbf{K}) \subseteq \mathcal{I}[\rho](\mathbf{K}') \quad \text{for every } \mathcal{D}\text{-cover } \mathbf{K}' \text{ of } L.$$

In general, there can be infinitely many optimal  $\mathcal{D}$ -covers for a given multiplicative rating map  $\rho$ . The key point is that there always exists at least one, provided that  $\mathcal{D}$  is a lattice. We state this property in the following lemma (proved in [32]).

**LEMMA 8.3.** Let  $\mathcal{D}$  be a lattice. For every language  $L$  and every multiplicative rating map  $\rho$ , there exists an optimal  $\mathcal{D}$ -cover of  $L$  for  $\rho$ .

Clearly, given a lattice  $\mathcal{D}$ , a language  $L$  and a multiplicative rating map  $\rho$ , all optimal  $\mathcal{D}$ -covers of  $L$  for  $\rho$  have the same  $\rho$ -imprint. Hence, this unique  $\rho$ -imprint is a *canonical* object for  $\mathcal{D}$ ,  $L$  and  $\rho$ . We call it the  *$\mathcal{D}$ -optimal  $\rho$ -imprint on  $L$*  and we write it  $\mathcal{I}_{\mathcal{D}}[L, \rho]$ :

$$\mathcal{I}_{\mathcal{D}}[L, \rho] = \mathcal{I}[\rho](\mathbf{K}) \quad \text{for any optimal } \mathcal{D}\text{-cover } \mathbf{K} \text{ of } L \text{ for } \rho.$$

An important special case is when  $L = A^*$ . In this case, we write  $\mathcal{I}_{\mathcal{D}}[\rho]$  for  $\mathcal{I}_{\mathcal{D}}[A^*, \rho]$ . Let us present a few properties of optimal imprints (all proved in [32]). First, we have the following useful fact, which is immediate from the definitions.

**FACT 8.4.** Let  $\rho$  be a multiplicative rating map and consider two languages  $H, L$  such that  $H \subseteq L$ . Then,  $\mathcal{I}_{\mathcal{D}}[H, \rho] \subseteq \mathcal{I}_{\mathcal{D}}[L, \rho]$ .

Additionally, the following lemma describes optimal imprints of a union of languages.

**LEMMA 8.5.** Let  $\mathcal{D}$  be a lattice and let  $\rho : 2^{A^*} \rightarrow R$  be a multiplicative rating map. Given two languages  $H, L$ , we have  $\mathcal{I}_{\mathcal{D}}[H \cup L, \rho] = \mathcal{I}_{\mathcal{D}}[H, \rho] \cup \mathcal{I}_{\mathcal{D}}[L, \rho]$

We complete Lemma 8.5 with a similar statement for language concatenation instead of union. Note that it requires more hypotheses on the class  $\mathcal{D}$ : we need closure under quotient.

**LEMMA 8.6.** Let  $\mathcal{D}$  be a positive prevariety and let  $\rho : 2^{A^*} \rightarrow R$  be a multiplicative rating map. Given two languages  $H, L \subseteq A^*$ , we have  $\mathcal{I}_{\mathcal{D}}[H, \rho] \cdot \mathcal{I}_{\mathcal{D}}[L, \rho] \subseteq \mathcal{I}_{\mathcal{D}}[HL, \rho]$ .

**Connection with covering.** We may now connect these definitions to the covering problem. The key idea is that solving  $\mathcal{D}$ -covering for a fixed class  $\mathcal{D}$  boils down to finding an algorithm that computes  $\mathcal{D}$ -optimal imprints from full rating maps given as inputs. In [32], two statements are presented. The first is simpler but it only applies to Boolean algebras, while the second is more involved and applies to all lattices. Since all classes investigated in the paper are Boolean algebras, we only present the first statement.

**PROPOSITION 8.7.** Let  $\mathcal{D}$  be a Boolean algebra. There exists an effective reduction from  $\mathcal{D}$ -covering to the following problem:

**Input:** A full rating map  $\rho : 2^{A^*} \rightarrow R$  and  $F \subseteq R$ .

**Question:** Is it true that  $\mathcal{I}_{\mathcal{D}}[\rho] \cap F = \emptyset$ ?

**PROOF SKETCH.** We briefly describe the reduction (we refer the reader to [32] for details). Consider an input pair  $(L_0, \{L_1, \dots, L_n\})$  for  $\mathcal{D}$ -covering. Since the languages  $L_i$  are regular, for every  $i \leq n$ , one can compute a morphism  $\alpha_i : A^* \rightarrow M_i$  into a finite monoid recognizing  $L_i$  together with the set  $F_i \subseteq M_i$  such that  $L_i = \alpha_i^{-1}(F_i)$ . Consider the associated full rating maps  $\rho_{\alpha_i} : 2^{A^*} \rightarrow 2^{M_i}$ . Moreover, let  $R$  be the idempotent semiring  $2^{M_0} \times \dots \times 2^{M_n}$  equipped with the componentwise addition and multiplication. We define a full rating map  $\rho : 2^{A^*} \rightarrow R$  by letting  $\rho(K) = (\rho_{\alpha_0}(K), \dots, \rho_{\alpha_n}(K))$  for every  $K \subseteq A^*$ . Finally, let  $F \subseteq R$  be the set of all tuples  $(X_0, \dots, X_n) \in R$  such that  $X_i \cap F_i \neq \emptyset$  for every  $i \leq n$ . One can now verify that  $(L_0, \{L_1, \dots, L_n\})$  is  $\mathcal{D}$ -coverable if and only if  $\mathcal{I}_{\mathcal{D}}[\rho] \cap F = \emptyset$ . Let us point out that this equivalence is only true when  $\mathcal{D}$  is a Boolean algebra. When  $\mathcal{D}$  is only a lattice, one has to handle the language  $L_0$  separately. ■

**Pointed optimal imprints.** In view of Proposition 8.7, given a Boolean algebra  $\mathcal{D}$ , an algorithm computing  $\mathcal{I}_{\mathcal{D}}[\rho]$  from a full rating map  $\rho$  yields a procedure for  $\mathcal{D}$ -covering. We consider the case where  $\mathcal{D} = UPol(C)$  for some finite prevariety  $C$ . We present a least fixpoint procedure for computing  $\mathcal{I}_{UPol(C)}[\rho]$ . Yet, implementing it requires considering an object which carries *more information* than  $\mathcal{I}_{UPol(C)}[\rho]$ , which we now define.

Let  $\mathcal{D}$  be a Boolean algebra,  $\eta : A^* \rightarrow N$  be a morphism into a finite monoid and  $\rho : 2^{A^*} \rightarrow R$  be a multiplicative rating map. The  $\eta$ -pointed  $\mathcal{D}$ -optimal  $\rho$ -imprint is defined as the following

set  $\mathcal{P}_{\mathcal{D}}[\eta, \rho] \subseteq N \times R$ :

$$\mathcal{P}_{\mathcal{D}}[\eta, \rho] = \{(t, r) \in N \times R \mid r \in \mathcal{I}_{\mathcal{D}}[\eta^{-1}(t), \rho]\}.$$

In view of this definition, we shall manipulate Cartesian products  $N \times R$  where  $N$  is a finite monoid and  $R$  is a finite idempotent semiring. In this context, it will be convenient to use the following functional notation. Given a subset  $S \subseteq N \times R$  and  $t \in N$ , we write  $S(t) \subseteq R$  for the set  $S(t) = \{r \in R \mid (t, r) \in S\}$ .

The set  $\mathcal{P}_{\mathcal{D}}[\eta, \rho]$  encodes the  $\mathcal{D}$ -optimal  $\rho$ -imprint on  $A^*$ , that is, the subset  $\mathcal{I}_{\mathcal{D}}[\rho]$  of  $R$ . Indeed, since  $\mathcal{I}_{\mathcal{D}}[\rho] = \mathcal{I}_{\mathcal{D}}[A^*, \rho]$ , we have the following immediate corollary of Lemma 8.5.

**COROLLARY 8.8.** Let  $\mathcal{D}$  be a Boolean algebra,  $\eta : A^* \rightarrow N$  be a morphism into a finite monoid and  $\rho : 2^{A^*} \rightarrow R$  be a multiplicative rating map. Then,

$$\mathcal{I}_{\mathcal{D}}[\rho] = \bigcup_{t \in N} \mathcal{I}_{\mathcal{D}}[\eta^{-1}(t), \rho] = \bigcup_{t \in N} \mathcal{P}_{\mathcal{D}}[\eta, \rho](t).$$

In the sequel, we consider the case where  $\mathcal{D} = \text{UPol}(C)$ , for some finite prevariety  $C$ . We present an algorithm for computing  $\mathcal{P}_{\text{UPol}(C)}[\eta_C, \rho] \subseteq N_C \times R$  from a full rating map  $\rho$ . Recall that  $\eta_C : A^* \rightarrow N_C$  is the canonical  $C$ -morphism, which is well-defined since  $C$  is a *finite* prevariety. This is actually the reason why the algorithm depends on the finiteness of  $C$ .

## 8.2 Characterization of $\text{UPol}(C)$ -optimal imprints

Let us first describe the property characterizing  $\text{UPol}(C)$ -optimal imprints for a finite prevariety  $C$ . Consider a morphism  $\eta : A^* \rightarrow N$  into a finite monoid (as explained above, in the statement,  $\eta$  will be the canonical  $C$ -morphism  $\eta_C$ ) and a multiplicative rating map  $\rho : 2^{A^*} \rightarrow R$ . We say that a subset  $S \subseteq N \times R$  is *UPol-saturated for  $\eta$  and  $\rho$*  if satisfies the four following properties:

1. *Trivial elements:* for every  $w \in A^*$ , we have  $(\eta(w), \rho(w)) \in S$ .
2. *Closure under downset:* we have  $\downarrow_R S = S$ .
3. *Closure under multiplication:* for every  $(s_1, r_1), (s_2, r_2) \in S$ , we have  $(s_1 s_2, r_1 r_2) \in S$ .
4. *UPol-closure:* if  $(e_1, f_1), (e_2, f_2) \in S$  are pairs of multiplicative idempotents and we have some  $s \in N$  such that  $e_1 \leq_{\mathcal{R}} s e_2$  and  $e_2 \leq_{\mathcal{L}} e_1 s$ , then  $(e_1 s e_2, f_1 \rho(\eta^{-1}(s)) f_2) \in S$ .

We are ready to state the main theorem of the section. Recall that when  $C$  is a finite prevariety, we write  $\eta_C : A^* \rightarrow N_C$  for the canonical  $C$ -morphism (see Section 2). We prove that for every multiplicative rating map  $\rho : 2^{A^*} \rightarrow R$ , the  $\eta_C$ -pointed  $\text{UPol}(C)$ -optimal  $\rho$ -imprint  $\mathcal{P}_{\text{UPol}(C)}[\eta_C, \rho]$  is the least *UPol-saturated* subset of  $N_C \times R$  (for inclusion) for  $\eta_C$  and  $\rho$ .

**THEOREM 8.9.** Let  $C$  be a finite prevariety and  $\rho : 2^{A^*} \rightarrow R$  be a multiplicative rating map. Then,  $\mathcal{P}_{\text{UPol}(C)}[\eta_C, \rho]$  is the least *UPol-saturated* subset of  $N_C \times R$  for  $\eta_C$  and  $\rho$ .

**REMARK 8.10.** When  $C$  is finite, Theorem 8.9 characterizes the least  $UPol$ -saturated subset for the canonical  $C$ -morphism and a multiplicative rating map. However, note that in order to obtain the decidability of  $UPol(C)$ -covering, we need the input to be finitely representable, *i.e.*, we need the rating map to be *full*. In this case, it is clear that one can use a least fixpoint procedure to compute the least  $UPol$ -saturated subset of  $N_C \times R$  from an input full rating map  $\rho : 2^{A^*} \rightarrow R$ . Therefore, Theorem 8.9 yields a procedure for computing  $\mathcal{P}_{UPol(C)}[\eta_C, \rho]$  from an input full rating map  $\rho : 2^{A^*} \rightarrow R$ . By Corollary 8.8, it follows that we may compute  $\mathcal{I}_{UPol(C)}[\rho] \subseteq R$  as well. Hence, Proposition 8.7 yields that  $UPol(C)$ -covering is decidable. This extends to  $UPol(C)$ -separation by Lemma 2.4.

From the above remark, we obtain the following corollary of Theorem 8.9.

**COROLLARY 8.11.** For every finite prevariety  $C$ ,  $UPol(C)$ -covering and  $UPol(C)$ -separation are decidable.

A key application of Corollary 8.11 is when  $C$  is the finite prevariety AT of alphabet testable languages: we get the decidability of covering and separation for  $UPol(AT) = UPol(BPol(ST))$  (we proved this equality in Lemma 4.25). Let us point out that this result was already known: it is proved in [32] using a specialized argument based on the logical characterization of  $UPol(AT)$  in terms of two-variable first-order logic (we present this logical characterization in Section 10). Actually, specializing the above generic algorithm yields exactly the procedure of [32].

We turn to the proof of Theorem 8.9. We present two independent statements, which correspond to soundness and completeness of the least fixpoint procedure that computes  $\mathcal{P}_{UPol(C)}[\eta_C, \rho]$ . We start with the former.

**PROPOSITION 8.12 (Soundness).** Let  $C$  be a finite prevariety and  $\rho : 2^{A^*} \rightarrow R$  be a multiplicative rating map. Then,  $\mathcal{P}_{UPol(C)}[\eta_C, \rho] \subseteq N_C \times R$  is  $UPol$ -saturated for  $\eta_C$  and  $\rho$ .

**PROOF.** Recall that  $UPol(C)$  is a prevariety by Theorem 4.24. There are four properties to verify. We start with the first three which are standard. For the trivial elements, consider  $w \in A^*$  and let  $\mathbf{K}$  be an optimal  $UPol(C)$ -cover of  $\eta_C^{-1}(\eta_C(w))$ . Since  $w \in \eta_C^{-1}(\eta_C(w))$ , there exists  $K \in \mathbf{K}$  such that  $w \in K$ . Thus,  $\rho(w) \leq \rho(K)$  which yields  $\rho(w) \in \mathcal{I}[\rho](\mathbf{K}) = \mathcal{I}_{UPol(C)}[\eta_C^{-1}(\eta_C(w)), \rho]$ . This implies that  $(\eta_C(w), \rho(w)) \in \mathcal{P}_{UPol(C)}[\eta_C, \rho]$ . Closure under downset is immediate by definition of imprints. Finally, for closure under multiplication, consider  $(s_1, r_1), (s_2, r_2) \in \mathcal{P}_{UPol(C)}[\eta_C, \rho]$ . We have  $r_i \in \mathcal{I}_{UPol(C)}[\eta_C^{-1}(s_i), \rho]$  for  $i = 1, 2$ . Since  $UPol(C)$  is a prevariety, Lemma 8.6 yields  $r_1 r_2 \in \mathcal{I}_{UPol(C)}[\eta_C^{-1}(s_1) \eta_C^{-1}(s_2), \rho]$ . Clearly,  $\eta_C^{-1}(s_1) \eta_C^{-1}(s_2) \subseteq \eta_C^{-1}(s_1 s_2)$ . Thus, Fact 8.4 yields  $r_1 r_2 \in \mathcal{I}_{UPol(C)}[\eta_C^{-1}(s_1 s_2), \rho]$ . By definition, this exactly says that  $(s_1 s_2, r_1 r_2) \in \mathcal{P}_{UPol(C)}[\eta_C, \rho]$ .

It remains to prove that  $\mathcal{P}_{UPol(C)}[\eta_C, \rho]$  satisfies  $UPol$ -closure. Consider two pairs of multiplicative idempotents  $(e_1, f_1), (e_2, f_2) \in \mathcal{P}_{UPol(C)}[\eta_C, \rho]$  and  $s \in N_C$  such that  $e_1 \leq_{\mathcal{R}} s e_2$  and

$e_2 \leq_{\mathcal{L}} e_1s$ . We prove that  $(e_1se_2, f_1\rho(\eta_C^{-1}(s))f_2) \in \mathcal{P}_{UPol(C)}[\eta_C, \rho]$ . By definition, we have to show that,

$$f_1\rho(\eta_C^{-1}(s))f_2 \in \mathcal{I}_{UPol(C)}[\eta_C^{-1}(e_1se_2), \rho].$$

By definition, this boils down to proving that given an arbitrary  $UPol(C)$ -cover  $\mathbf{K}$  of  $\eta_C^{-1}(e_1se_2)$ , we have  $f_1\rho(\eta_C^{-1}(s))f_2 \in \mathcal{I}[\rho](\mathbf{K})$ . We fix  $\mathbf{K}$  for the proof. Proposition 2.13 yields a  $UPol(C)$ -morphism  $\alpha : A^* \rightarrow M$  recognizing every language  $K \in \mathbf{K}$ . Let  $k = \omega(M)$ .

Let  $i \in \{1, 2\}$ . By hypothesis, we know that  $(e_i, f_i) \in \mathcal{P}_{UPol(C)}[\eta_C, \rho]$ . By definition, this means that  $f_i \in \mathcal{I}_{UPol(C)}[\eta_C^{-1}(e_i), \rho]$ . Clearly, the set  $\{\alpha^{-1}(t) \mid t \in M \text{ and } \alpha^{-1}(t) \cap \eta_C^{-1}(e_i) \neq \emptyset\}$  is a  $UPol(C)$ -cover of  $\eta_C^{-1}(e_i)$ . Therefore, since  $f_i \in \mathcal{I}_{UPol(C)}[\eta_C^{-1}(e_i), \rho]$ , we get  $t_i \in M$  such that  $\eta_C^{-1}(e_i) \cap \alpha^{-1}(t_i) \neq \emptyset$  and  $f_i \leq \rho(\alpha^{-1}(t_i))$ . We fix  $w_i \in \eta_C^{-1}(e_i) \cap \alpha^{-1}(t_i)$  for the proof. Moreover, we let  $u \in \eta_C^{-1}(s)$ . We define,

$$w = w_1^k u w_2^k.$$

Since  $e_1, e_2 \in N_C$  are idempotents, we have  $\eta_C(w) = e_1se_2$ . Since  $\mathbf{K}$  is a cover of  $\eta_C^{-1}(e_1se_2)$ , we get  $K \in \mathbf{K}$  such that  $w \in K$ . We prove that,

$$(\alpha^{-1}(t_1))^k \eta_C^{-1}(s) (\alpha^{-1}(t_2))^k \subseteq K. \quad (11)$$

Let us first explain how to use (11) to conclude the proof. Since  $f_i \leq \rho(\alpha^{-1}(t_i))$  for  $i = 1, 2$ , the inclusion given by (11) implies that  $f_1^k \rho(\eta_C^{-1}(s)) f_2^k \leq \rho(K)$ . Moreover, since  $f_1, f_2 \in R$  are multiplicative idempotents, this yields  $f_1 \rho(\eta_C^{-1}(s)) f_2 \leq \rho(K)$ . Finally, since  $K \in \mathbf{K}$ , we get  $f_1 \rho(\eta_C^{-1}(s)) f_2 \in \mathcal{I}[\rho](\mathbf{K})$ , as desired.

It remains to prove that (11) holds. We fix  $w' \in (\alpha^{-1}(t_1))^k \eta_C^{-1}(s) (\alpha^{-1}(t_2))^k$  for the proof and show that  $w' \in K$ . Since  $K$  is recognized by  $\alpha$  and  $w \in K$ , it suffices to prove that  $\alpha(w') = \alpha(w)$ . For  $i = 1, 2$ , we write  $g_i = t_i^\omega$ . By definition of  $w$  and of  $k = \omega(M)$ , we have  $\alpha(w) = g_1 \alpha(u) g_2$ . By hypothesis on  $w'$ , there exists  $u' \in \eta_C^{-1}(s)$  such that  $\alpha(w') = g_1 \alpha(u') g_2$ . Hence, it remains to show that  $g_1 \alpha(u) g_2 = g_1 \alpha(u') g_2$ . First, we use our hypothesis on  $e_1, e_2, s \in N_C$ : since  $e_1 \leq_{\mathcal{R}} se_2$  and  $e_2 \leq_{\mathcal{L}} e_1s$ , there exist  $q_1, q_2 \in N_C$  such that  $e_1 = se_2q_1$  and  $e_2 = q_2e_1s$ . Let  $p = q_2e_1se_2q_1 \in N_C$ . Since  $e_1$  and  $e_2$  are idempotents, we have,

$$\begin{aligned} e_1 &= se_2q_1 = se_2e_2e_2q_1 = se_2q_2e_1se_2q_1 = se_2p, \\ e_2 &= q_2e_1s = q_2e_1e_1e_1s = q_2e_1se_2q_1e_1s = pe_1s. \end{aligned}$$

Let  $v \in \eta_C^{-1}(p)$ . Since  $\eta_C(w_i^k) = e_i$  for  $i = 1, 2$  and  $\eta_C(u) = \eta_C(u') = s$ , the above yields the equalities  $\eta_C(w_1^k) = \eta_C(u'w_2^kv)$  and  $\eta_C(w_2^k) = \eta_C(vw_1^ku)$ . Since  $\eta_C$  is the canonical  $C$ -morphism and  $\alpha(w_i^k) = g_i$  for  $i = 1, 2$ , one obtain, using Lemma 2.16 and Lemma 5.19, that  $g_1 \sim_C \alpha(u')g_2\alpha(v)$  and  $g_2 \sim_C \alpha(v)g_1\alpha(u)$ . Finally, since  $\alpha$  is a  $UPol(C)$ -morphism, Theorem 7.2 implies that it satisfies (10). Since  $g_1$  and  $g_2$  are idempotents, this yields:  $g_1 = g_1\alpha(u')g_2\alpha(v)g_1$

and  $g_2 = g_2\alpha(v)g_1\alpha(u)g_2$ . It follows that,

$$g_1\alpha(u)g_2 = g_1\alpha(u')g_2\alpha(v)g_1\alpha(u)g_2 = g_1\alpha(u')g_2.$$

This concludes the proof. ■

We turn to the completeness direction in Theorem 8.9.

**PROPOSITION 8.13 (Completeness).** Let  $C$  be a prevariety,  $\eta : A^* \rightarrow N$  be a  $C$ -morphism,  $\rho : 2^{A^*} \rightarrow R$  be a multiplicative rating map and  $S \subseteq N \times R$  be  $UPol$ -saturated for  $\eta$  and  $\rho$ . For every  $t \in N$ , there exists a  $UPol(C)$ -cover  $\mathbf{K}_t$  of  $\eta^{-1}(t)$  such that  $\mathcal{I}[\rho](\mathbf{K}_t) \subseteq S(t)$ .

**PROOF.** We fix  $C$ ,  $\eta$ ,  $\rho$  and  $S$  as in the statement. Since  $S$  is  $UPol$ -saturated, it is closed under multiplication, which implies that  $S$  is a monoid for the componentwise multiplication (the identity element is the trivial element  $(1_N, 1_R) = (\eta(\varepsilon), \rho(\varepsilon))$ ). The proposition is a corollary of the following lemma, which we prove by induction.

**LEMMA 8.14.** Let  $(x, p), (y, q) \in S$  and  $t \in N$ . There exists a  $UPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(t)$  such that  $(xty, p\rho(K)q) \in S$  for every  $K \in \mathbf{K}$ .

We first apply the lemma to complete the proof of Proposition 8.13. We apply it for  $(x, p) = (y, q) = (1_N, 1_R) \in S$ . For every  $t \in N$ , this yields a  $UPol(C)$ -partition  $\mathbf{K}_t$  of  $\eta^{-1}(t)$  such that  $(t, \rho(K)) \in S$  for every  $K \in \mathbf{K}_t$ . Since  $S$  is closed under downset, the fact that  $(t, \rho(K)) \in S$  for every  $K \in \mathbf{K}_t$  implies that  $\mathcal{I}[\rho](\mathbf{K}_t) \subseteq S(t)$ , as desired.

It remains to prove Lemma 8.14. The proof is reminiscent of that of Lemma 7.6. Let  $(x, p), (y, q) \in S$  and  $t \in N$ . We build a  $UPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(t)$  such that  $(xty, p\rho(K)q) \in S$  for every  $K \in \mathbf{K}$  by induction on the following three parameters listed by order of importance (they depend on the Green relations of the finite monoids  $N$  and  $S$ ).

1. The *rank of  $xty \in N$* : the number of elements  $t' \in N$  such that  $xty \leq_{\mathcal{J}} t'$ .
2. The  *$\mathcal{R}$ -index of  $(x, p)$* : the number of pairs  $(x', p') \in S$  such that  $(x', p') \leq_{\mathcal{R}} (x, p)$ .
3. The  *$\mathcal{L}$ -index of  $(y, q)$* : the number of pairs  $(y', q') \in S$  such that  $(y', q') \leq_{\mathcal{L}} (y, q)$ .

We distinguish three cases depending on whether  $xty \mathcal{J} t$  and on which of the two following properties of  $t$ ,  $(x, p)$  and  $(y, q)$  are fulfilled:

- $(x, p)$  is *right  $t$ -stable* when there is  $(z, r) \in S$  such that  $(xz, pr) \mathcal{R} (x, p)$  and  $z \mathcal{R} t$ .
- $(y, q)$  is *left  $t$ -stable* when there is  $(z, r) \in S$  such that  $(zy, rq) \mathcal{L} (y, q)$  and  $z \mathcal{L} t$ .

In the base case, we assume that all three properties holds and conclude directly. Otherwise, we consider two distinct inductive cases. In the first one, we assume that  $xty <_{\mathcal{J}} t$  and in the second one, that either  $(x, p)$  is not right  $t$ -stable or  $(y, q)$  is not left  $t$ -stable.

**Base case:**  $xty \mathcal{J} t$ ,  $(x, p)$  is right  $t$ -stable and  $(y, q)$  is left  $t$ -stable. We let  $\mathbf{K} = \{\eta^{-1}(t)\}$ . Clearly, this is a  $UPol(C)$ -partition of  $\eta^{-1}(t)$  since  $\eta$  is a  $C$ -morphism. Hence, it remains to prove that  $(xty, p\rho(\eta^{-1}(t))q) \in S$ . We first use our hypothesis to prove the following fact.

**FACT 8.15.** There exist two pairs of multiplicative idempotents  $(e_1, f_1), (e_2, f_2) \in S$  satisfying the following conditions:  $(xe_1, pf_1) = (x, p)$ ,  $(e_2y, f_2q) = (y, q)$ ,  $e_1 \leq_{\mathcal{R}} te_2$  and  $e_2 \leq_{\mathcal{L}} e_1t$ .

**PROOF.** We exhibit  $(e_1, f_1)$  using the right  $t$ -stability of  $(x, p)$ . Since  $(x, p)$  is right  $t$ -stable, there exists  $(z, r) \in S$  such that  $(xz, pr) \mathcal{R} (x, p)$  and  $z \mathcal{R} t$ . Hence, we obtain  $(z', r') \in S$  such that  $(x, p) = (xzz', prr')$ . Let  $k = \omega(S) \geq 1$ . Then,  $(e_1, f_1) = ((zz')^k, (rr')^k)$  is a pair of multiplicative idempotents. It is now immediate that  $(xe_1, pf_1) = (x, p)$ . The existence of  $(e_2, f_2)$  is proved symmetrically using the left  $t$ -stability of  $(y, q)$ . We now show that  $e_1 \leq_{\mathcal{R}} te_2$  (the proof that  $e_2 \leq_{\mathcal{L}} e_1t$  is symmetrical and left to the reader). Since  $z \mathcal{R} t$ , it is clear that  $e_1 = (zz')^k \leq_{\mathcal{R}} t$ . Moreover, since  $xy = xe_1te_2y \mathcal{J} t$  by hypothesis, we have  $te_2 \mathcal{J} t$ . Since it is clear that  $te_2 \leq_{\mathcal{R}} t$ , Lemma 2.1 yields  $te_2 \mathcal{R} t$ . Altogether, we get  $e_1 \leq_{\mathcal{R}} te_2$ . ■

We let  $(e_1, f_1), (e_2, f_2) \in S$  be given by Fact 8.15. Since  $(e_1, f_1), (e_2, f_2) \in S$  are multiplicative idempotents such that  $e_1 \leq_{\mathcal{R}} te_2$  and  $e_2 \leq_{\mathcal{L}} e_1t$ , and  $S$  is  $UPol$ -saturated, we get from  $UPol$ -closure that  $(e_1te_2, f_1\rho(\eta^{-1}(t))f_2) \in S$ . Finally, since we have  $(x, p), (y, q) \in S$ ,  $(xe_1, pf_1) = (x, p)$  and  $(e_2y, f_2q) = (y, q)$ , we obtain from closure under multiplication, that  $(x, p)(e_1te_2, f_1\rho(\eta^{-1}(t))f_2)(y, q) = (xty, p\rho(\eta^{-1}(t))q) \in S$ , which concludes this case.

**First inductive case:**  $xty <_{\mathcal{J}} t$ . In this case, the rank of  $t = 1_N t 1_N$  is strictly smaller than the one of  $xty$ . Consequently, induction on our first parameter in Lemma 8.14 (applied for  $(x, q) = (y, r) = (1_N, 1_R) \in S$ ) yields a  $UPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(t)$  such that  $(t, \rho(K)) \in S$  for every  $K \in \mathbf{K}$ . Since  $(x, q), (y, r) \in S$  and  $S$  is closed under multiplication, it then follows that  $(xty, q\rho(K)r) \in S$  for every  $K \in \mathbf{K}$ , concluding this case.

**Second inductive case:** either  $(x, p)$  is not right  $t$ -stable or  $(y, q)$  is not left  $t$ -stable. There are two symmetrical cases depending on which property holds. We handle the case where  $(y, q)$  is not left  $t$ -stable and leave the other, which follows by symmetry, to the reader. Let  $T$  be the set of all triples  $(t_1, a, t_2) \in N \times A \times N$  such that  $t_1\eta(a)t_2 = t$  and  $t \mathcal{L} \eta(a)t_2 <_{\mathcal{L}} t_2$ . In the next fact, we use induction to build  $UPol(C)$ -partitions of  $\eta^{-1}(t_2)$  and  $\eta^{-1}(t_1)$  for every triple  $(t_1, a, t_2) \in T$ . We shall then combine them to construct the desired  $UPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(t)$ .

**FACT 8.16.** Consider a triple  $(t_1, a, t_2) \in T$ . There exists a  $UPol(C)$ -partition  $\mathbf{V}_{t_2}$  of  $\eta^{-1}(t_2)$  such that  $(t_2, \rho(V)) \in S$  for every  $V \in \mathbf{V}_{t_2}$ . Moreover, for every  $V \in \mathbf{V}_{t_2}$ , there exists a  $UPol(C)$ -partition  $\mathbf{U}_{(t_1, a, t_2), V}$  of  $\eta^{-1}(t_1)$  such that  $(xty, p\rho(UaV)q) \in S$  for every  $U \in \mathbf{U}_{(t_1, a, t_2), V}$ .

**PROOF.** By definition of  $T$ , we know that  $t <_{\mathcal{L}} t_2$ . This implies  $t <_{\mathcal{J}} t_2$  by Lemma 2.1 and since  $xty \leq_{\mathcal{J}} t$ , we get  $xty <_{\mathcal{J}} t_2$ . Hence, the rank of  $t_2 = 1_N t_2 1_N$  is strictly smaller than the one of  $xty$ . Therefore, by induction on our first and main parameter in Lemma 8.14 (applied for  $(x, p) = (y, q) = (1_N, 1_R)$ ) we obtain a  $UPol(C)$ -partition  $\mathbf{V}_{t_2}$  of  $\eta^{-1}(t_2)$  such that  $(t_2, \rho(V)) \in S$  for every  $V \in \mathbf{V}_{t_2}$ .

We now fix  $V \in \mathbf{V}_{t_2}$  and build  $\mathbf{U}_{(t_1, a, t_2), V}$ . Since  $(t_2, \rho(V)) \in S$  by definition of  $\mathbf{V}_{t_2}$  and  $(\eta(a), \rho(a)) \in S$  (this is a trivial element), we have  $(\eta(a)t_2, \rho(aV)) \in S$ . Hence, since  $t \mathcal{L} \eta(a)t_2$  by definition of  $T$ , the hypothesis that  $(y, q)$  is not left  $t$ -stable yields  $(\eta(a)t_2y, \rho(aV)q) <_{\mathcal{L}} (y, q)$ . Consequently, the  $\mathcal{L}$ -index of  $(\eta(a)t_2y, \rho(aV)q)$  is strictly smaller than the one of  $(y, q)$ . We may apply induction to  $(x, p)$ ,  $(\eta(a)t_2y, \rho(aV)q)$  and  $t_1$ . Indeed, while the third parameter has decreased, the first has not increased since  $xt_1\eta(a)t_2y = xty$  by definition of  $T$ . The second one has not increased as well since we did not change the first pair  $(x, p)$ . Hence, induction on our third parameter in Lemma 8.14 applied when  $(y, q)$  has been replaced by  $(\eta(a)t_2y, \rho(aV)q)$  and  $t$  by  $t_1$  yields a  $UPol(C)$ -partition  $\mathbf{U}_{(t_1, a, t_2), V}$  of  $\eta^{-1}(t_1)$  such that  $(xt_1\eta(a)t_2y, p\rho(U)\rho(aV)q) \in S$  for every  $U \in \mathbf{U}_{(t_1, a, t_2), V}$ . Finally, since  $xt_1\eta(a)t_2y = xty$ , we have  $(xty, p\rho(UaV)q) \in S$  for every  $U \in \mathbf{U}_{(t_1, a, t_2), V}$ . ■

We are ready to construct our  $UPol(C)$ -partition  $\mathbf{K}$  of  $\eta^{-1}(t)$ . We define,

$$\mathbf{K} = \bigcup_{(t_1, a, t_2) \in T} \{UaV \mid V \in \mathbf{V}_{t_2} \text{ and } U \in \mathbf{U}_{(t_1, a, t_2), V}\}.$$

It is immediate from the definition and Fact 8.16 that  $(xty, p\rho(K)q) \in S$  for every  $K \in \mathbf{K}$ . It remains to verify that  $\mathbf{K}$  is a  $UPol(C)$ -partition of  $\eta^{-1}(t)$ . We first prove that it is a partition of  $\eta^{-1}(t)$ : let  $w \in \eta^{-1}(t)$ , we show that there is a *unique*  $K \in \mathbf{K}$  such that  $w \in K$ . Let  $v' \in A^*$  be the least suffix of  $w$  such that  $t = \eta(w) \mathcal{L} \eta(v')$  and let  $u \in A^*$  such that  $w = uv'$ . Observe that  $v' \neq \varepsilon$  (otherwise, we would have  $t \mathcal{L} 1_N$ , contradicting the hypothesis that  $(y, q)$  is not left  $t$ -stable, since  $(1_Ny, 1_Rq) \mathcal{L} (y, q)$ ). Hence, there exists  $v \in A^*$  and  $a \in A$  such that  $v' = av$ . Let  $t_1 = \eta(u)$  and  $t_2 = \eta(v)$ . By definition  $t_1\eta(a)t_2 = \eta(uav) = \eta(w) = t$ . Moreover,  $t \mathcal{L} \eta(v') = \eta(a)t_2 <_{\mathcal{L}} \eta(v)$  by definition of  $v'$  as the least prefix of  $w$  such that  $t = \eta(w) \mathcal{L} \eta(v')$ . Hence,  $(t_1, a, t_2) \in T$  and we may consider the partition  $\mathbf{V}_{t_2}$  of  $\eta^{-1}(t_2)$ . In particular, there exists  $V \in \mathbf{V}_{t_2}$  such that  $v \in V$  since  $t_2 = \eta(v)$ . Moreover, since  $\mathbf{U}_{(t_1, a, t_2), V}$  is a partition of  $\eta^{-1}(t_1)$  and  $t_1 = \eta(u)$ , there exists  $U \in \mathbf{U}_{(t_1, a, t_2), V}$  such that  $u \in U$ . Hence,  $w = uav \in UaV$  and  $UaV \in \mathbf{K}$  is the unique language of  $\mathbf{K}$  containing  $w$  by definition.

Finally, let us verify that for every  $K \in \mathbf{K}$ , we have  $K \subseteq \eta^{-1}(t)$  and  $K \in UPol(C)$  which completes the proof that  $\mathbf{K}$  is a  $UPol(C)$ -partition of  $\eta^{-1}(t)$ . Let  $K \in \mathbf{K}$ . By definition,  $K = UaV$  where  $V \in \mathbf{V}_{t_2}$  and  $U \in \mathbf{U}_{(t_1, a, t_2), V}$  for some  $(t_1, a, t_2) \in T$ . Since  $\mathbf{V}_{t_2}$  and  $\mathbf{U}_{(t_1, a, t_2), V}$  are  $UPol(C)$ -partitions of  $\eta^{-1}(t_2)$  and  $\eta^{-1}(t_1)$  respectively, we know that  $U, V \in UPol(C)$ ,  $U \subseteq \eta^{-1}(t_1)$  and  $V \subseteq \eta^{-1}(t_2)$ . It follows that  $K \subseteq \eta^{-1}(t_1)a\eta^{-1}(t_2)$ . Hence, since  $t_1\eta(a)t_2 = t$  by definition of  $T$ , we get  $K \subseteq \eta^{-1}(t)$ . Moreover,  $U, V \in UPol(C)$  and  $UaV$  is right deterministic by Lemma 4.15 since  $V \subseteq \eta^{-1}(t_2)$  and  $t_2\eta(a) <_{\mathcal{L}} t_2$  (as  $(t_1, a, t_2) \in T$ ). This concludes the proof. ■

We are now ready to prove Theorem 8.9.

**PROOF OF THEOREM 8.9.** Let  $C$  be a finite prevariety and  $\rho : 2^{A^*} \rightarrow R$  be a multiplicative rating map. By Proposition 8.12,  $\mathcal{I}_{UPol(C)}[\rho] \subseteq N_C \times R$  is  $UPol$ -saturated for  $\eta_C$  and  $\rho$ . It remains

to show that it is the least such set. Thus, let  $S \subseteq N_C \times R$  be  $UPol$ -saturated for  $\eta_C$  and  $\rho$ . We show that  $\mathcal{P}_{UPol(C)}[\eta_C, \rho] \subseteq S$ . Clearly, it suffices to show that  $\mathcal{P}_{UPol(C)}[\eta_C, \rho](t) \subseteq S(t)$  for every  $t \in N_C$ . We fix  $t$  for the proof.

Since  $\eta_C$  is a  $C$ -morphism, Proposition 8.13 yields a  $UPol(C)$ -cover  $\mathbf{K}$  of  $\eta_C^{-1}(t)$  such that  $\mathcal{I}[\rho](\mathbf{K}_t) \subseteq S(t)$ . Moreover, recall that  $\mathcal{P}_{UPol(C)}[\eta_C, \rho](t) = \mathcal{I}_{UPol(C)}[\eta_C^{-1}(t), \rho]$ . Therefore, since  $\mathbf{K}$  is a  $UPol(C)$ -cover of  $\eta_C^{-1}(t)$ , we get  $\mathcal{P}_{UPol(C)}[\eta_C, \rho](t) \subseteq \mathcal{I}[\rho](\mathbf{K})$ . Altogether, we obtain  $\mathcal{P}_{UPol(C)}[\eta_C, \rho](t) \subseteq S(t)$ , as desired. ■

## 9. Unary temporal logic

We introduce *unary temporal logic*. We use a definition that generalizes the standard one: for each class  $C$ , we define a particular variant of unary temporal logic that we denote by  $TL[C]$  and associate a class  $TL(C)$  to it. The standard definitions of unary temporal logic found in the literature correspond either to  $TL(ST)$  or to  $TL(ST^+)$ .

We prove two key results in the section. First, we establish a connection with two-variable first-order logic. For every Boolean algebra  $C$ , we prove that  $TL(C) = FO^2(\mathbb{I}_C)$ . This result generalizes a well-known theorem by Etessami, Vardi and Wilke [9]. We also compare these logical classes to those built with unambiguous polynomial closure. More precisely, we prove the given an arbitrary prevariety  $C$ , we have the inclusion  $UPol(BPol(C)) \subseteq TL(C)$ . While it is strict in general, we shall prove in the next section that when  $C$  is a prevariety of group languages or a well-suited extension thereof, the converse inclusion holds as well.

### 9.1 Definition and properties

We actually define two distinct sets of temporal formulas, which we denote by  $TLX$  and  $TL$ . Then, we explain how these sets can be restricted depending on some class of languages  $C$ . This yields two new classes built from  $C$ , which we write  $TLX(C)$  and  $TL(C)$ . Let us first define the  $TLX$  formulas, which are more general.

A  $TLX$  formula is built from atomic formulas using Boolean connectives and temporal operators. The atomic formulas are  $\top$ ,  $\perp$ ,  $min$ ,  $max$  and “ $a$ ” for every letter  $a \in A$ . All Boolean connectives are allowed: if  $\psi_1$  and  $\psi_2$  are  $TLX$  formulas, then so are  $(\psi_1 \vee \psi_2)$ ,  $(\psi_1 \wedge \psi_2)$  and  $(\neg\psi_1)$ . There are two kinds of *temporal operators*, which are both unary. First, one may use  $X$  and  $Y$ : if  $\psi$  is a  $TLX$  formula, then so are  $(X \psi)$  and  $(Y \psi)$ . Moreover, we associate *two temporal operators* to every language  $L \subseteq A^*$ , which we write  $F_L$  and  $P_L$ : if  $\psi$  is a  $TLX$  formula, then so are  $(F_L \psi)$  and  $(P_L \psi)$ . For the sake of improved readability, we omit parentheses when there is no ambiguity. Finally, we define  $TL$  as a syntactical restriction: a  $TL$  formula is a  $TLX$  formula that does *not* contain  $X$ , nor  $Y$ .

We now turn to the semantics. Since  $TL$  formulas are particular  $TLX$  formulas, it suffices to define the semantics of the latter. Evaluating a  $TLX$  formula  $\varphi$  requires a word  $w \in A^*$  and

a position  $i \in \text{Pos}(w)$ . We use structural induction on  $\varphi$  to define what it means for  $(w, i)$  to satisfy  $\varphi$ . We denote this property by  $w, i \models \varphi$ :

- **Atomic formulas:**  $w, i \models \top$  always holds and  $w, i \models \perp$  never holds. Additionally, for every  $\ell \in A \cup \{\min, \max\}$ ,  $w, i \models \ell$  holds when  $\ell = w[i]$ .
- **Disjunction:**  $w, i \models \psi_1 \vee \psi_2$  when  $w, i \models \psi_1$  or  $w, i \models \psi_2$ .
- **Conjunction:**  $w, i \models \psi_1 \wedge \psi_2$  when  $w, i \models \psi_1$  and  $w, i \models \psi_2$ .
- **Negation:**  $w, i \models \neg\psi$  when  $w, i \models \psi$  does not hold.
- **Next:**  $w, i \models X \varphi$  when  $i + 1$  is a position of  $w$  and  $w, i + 1 \models \varphi$ .
- **Preceding:**  $w, i \models Y \varphi$  when  $i - 1$  is a position of  $w$  and  $w, i - 1 \models \varphi$ .
- **Finally:** for  $L \subseteq A^*$ , we let  $w, i \models F_L \psi$  when there exists a position  $j > i$  of  $w$  such that  $w, j \models \psi$  and  $w(i, j) \in L$ .
- **Previously:** for  $L \subseteq A^*$ , we let  $w, i \models P_L \psi$  when there exists a position  $j < i$  of  $w$  such that  $w, j \models \psi$  and  $w(j, i) \in L$ .

When no distinguished position is specified, we evaluate formulas at the *leftmost unlabeled position*. More precisely, given a TLX formula  $\varphi$  and a word  $w \in A^*$ , we write  $w \models \varphi$  and say that  $w$  satisfies  $\varphi$  if and only if  $w, 0 \models \varphi$ . Finally, the *language defined by  $\varphi$*  is  $L(\varphi) = \{w \in A^* \mid w \models \varphi\}$ . Of course, considering *all* formulas is not really interesting, as every language  $L$  is defined by “ $F_L \max$ ”.

Consider a class  $C$ . We first explain how  $C$  can be used to restrict the sets of TL and TLX formulas. A TL[ $C$ ] (resp. TLX[ $C$ ]) formula is a TL (resp. TLX) formula  $\varphi$  such that every temporal operator  $F_L$  or  $P_L$  occurring in  $\varphi$  satisfies  $L \in C$ . With this definition in hand, we associate two classes to  $C$ , which we denote by TL( $C$ ) and TLX( $C$ ). They consist of all languages that can be defined by a TL[ $C$ ] and a TLX[ $C$ ] formula, respectively. Observe that by definition, Boolean connectives can be used freely in TL and TLX formulas, Hence, it is immediate that the two associated classes are Boolean algebras. They both contain  $C$ : a language  $L \in C$  is defined by the TL[ $C$ ] formula “ $F_L \max$ ”. One may also prove that if  $C$  is a prevariety, then so are TL( $C$ ) and TLX( $C$ ). Yet, we shall not use this property.

**REMARK 9.1.** It is immediate that for an arbitrary TLX formula  $\varphi$ ,  $w \in A^*$  and  $i \in \text{Pos}(w)$ , we have  $w, i \models F_{A^*} \varphi$  (resp.  $w, i \models P_{A^*} \varphi$ ) if and only if there exists  $j \in \text{Pos}(w)$  such that  $i < j$  (resp.  $j < i$ ) and  $w, j \models \varphi$ . Hence,  $F_{A^*}$  and  $P_{A^*}$  correspond to standard operators in unary temporal logic, usually denoted by  $F$  and  $F^{-1}$ . In particular for the input class  $\text{ST} = \{\emptyset, A^*\}$ , the classes TL(ST) and TLX(ST) correspond to the standard variants of unary temporal logic, which are often denoted by  $F + F^{-1}$  and  $F + X + F^{-1} + X^{-1}$ . For the sake of consistency, we write  $P$  (for “past”) instead of  $F^{-1}$  and  $Y$  (for “yesterday”) instead of  $X^{-1}$ .

We now prove that when  $C$  is a prevariety, the operator  $C \mapsto \text{TL}(C)$  is “more fundamental” than  $C \mapsto \text{TLX}(C)$ . More precisely, we show that if  $C$  is a prevariety, then  $\text{TLX}(C) = \text{TL}(C^+)$ .

The takeaway is that while considering  $\text{TL}(C)$  and  $\text{TLX}(C)$  independently is natural when presenting statements, it suffices to consider  $\text{TL}(C)$  in proof arguments.

**LEMMA 9.2.** Let  $C$  be a prevariety. Then,  $\text{TLX}(C) = \text{TL}(C^+)$ .

**PROOF.** We start with  $\text{TLX}(C) \subseteq \text{TL}(C^+)$ . By definition, it suffices to show that  $X$  and  $Y$  can be expressed with the operators available in  $\text{TL}[C^+]$  formulas. This is immediate since  $\{\varepsilon\} \in C^+$  and the operators  $F_{\{\varepsilon\}}$  and  $P_{\{\varepsilon\}}$  have the same semantics as  $X$  and  $Y^{-1}$ .

We turn to the inclusion  $\text{TL}(C^+) \subseteq \text{TLX}(C)$ . Given an arbitrary  $\text{TL}[C^+]$  formula  $\varphi$ , we explain how to inductively construct an equivalent  $\text{TLX}[C]$  formula  $\langle \varphi \rangle$  (i.e., for every word  $w \in A^*$  and every position  $i \in \text{Pos}(w)$ , we have  $w, i \models \varphi \Leftrightarrow w, i \models \langle \varphi \rangle$ ). By definition, this yields  $\text{TL}(C^+) \subseteq \text{TLX}(C)$ , as desired. If  $\varphi$  is atomic, it is already a  $\text{TLX}[C]$  formula and we may define  $\langle \varphi \rangle := \varphi$ . Boolean combinations are handled in the natural way. By definition of  $\text{TL}[C^+]$  formulas, it remains to handle the case where  $\varphi$  is of the form  $F_L \psi$  or  $P_L \psi$  for some  $L \in C^+$ . By symmetry, we only consider the case where  $\varphi := F_L \psi$ . By definition of  $C^+$ , there exists  $K \in C$  such that either  $L = \{\varepsilon\} \cup K$  or  $L = A^+ \cap K$ . In the former case, it suffices to define  $\langle \varphi \rangle := (X \langle \psi \rangle) \vee (F_K \langle \psi \rangle)$ , which is a  $\text{TLX}[C]$  formula by definition. Otherwise, we have  $L = A^+ \cap K$ . Since  $C$  is a prevariety, we know that  $a^{-1}K \in C$  for every  $a \in A$ . Hence, we may define  $\langle \varphi \rangle$  as the following  $\text{TLX}[C]$  formula:

$$\langle \varphi \rangle := X \left( \bigvee_{a \in A} (a \wedge F_{a^{-1}K} \langle \psi \rangle) \right).$$

This concludes the proof. ■

We complete the definition with a property of the classes  $\text{TL}(C)$  (when  $C$  is a prevariety) which we shall need in Section 10 to establish the correspondence with unambiguous polynomial closure. This involves quite a bit of work as we require some machinery to present it. First, we define equivalence relations that we shall use to formulate the property.

**Canonical equivalences.** We start with some terminology needed for the definition. Given a morphism  $\eta : A^* \rightarrow N$  into a finite monoid  $N$ , a  $\text{TL}[\eta]$  formula is a  $\text{TL}$  formula  $\varphi$  such that for every operator  $F_L$  or  $P_L$  occurring in  $\varphi$ , the language  $L \subseteq A^*$  is recognized by  $\eta$ . The following simple fact connects this notion to the classes  $\text{TL}(C)$ .

**FACT 9.3.** Let  $C$  be a prevariety. For every  $\text{TL}[C]$  formula  $\varphi$ , there exists a  $C$ -morphism  $\eta : A^* \rightarrow N$  such that  $\varphi$  is a  $\text{TL}[\eta]$  formula.

**PROOF.** Let  $\mathbf{L}$  be the finite set consisting of all languages  $L \subseteq A^*$  such that either  $F_L$  or  $P_L$  occurs in  $\varphi$ . Since  $\varphi$  is a  $\text{TL}[C]$  formula, we know that every  $L \in \mathbf{L}$  belongs to  $C$ . By hypothesis on  $C$ , it follows from Proposition 2.13 that there exists a  $C$ -morphism  $\eta : A^* \rightarrow N$  recognizing every  $L \in \mathbf{L}$ . By definition,  $\varphi$  is a  $\text{TL}[\eta]$  formula. ■

We now associate a number called *rank* to every TL formula  $\varphi$  (this is a standard notion in unary temporal logic). As expected, the rank of  $\varphi$  is defined as the length of the longest sequence of nested temporal operators within its parse tree. More precisely,

- Any atomic formula has rank 0.
- The rank of  $\neg\varphi$  is the same as the rank of  $\varphi$ .
- The rank of  $\varphi \vee \psi$  and  $\varphi \wedge \psi$  is the maximum between the ranks of  $\varphi$  and  $\psi$ .
- For every language  $L \subseteq A^*$ , the rank of  $F_L \varphi$  and  $P_L \varphi$  is the rank of  $\varphi$  plus 1.

Two TL formulas  $\varphi$  and  $\psi$  are *equivalent* if they have the same semantics. That is, for every  $w \in A^*$  and every position  $i \in \text{Pos}(w)$ , we have  $w, i \models \varphi \Leftrightarrow w, i \models \psi$ . The following key lemma is immediate from a simple induction on the rank of TL formulas.

**LEMMA 9.4.** Let  $\eta : A^* \rightarrow N$  be a morphism into a finite monoid and let  $k \in \mathbb{N}$ . There are only finitely many non-equivalent  $\text{TL}[\eta]$  formulas with rank at most  $k$ .

We may now define the equivalences associated to TL. They relate pairs  $(w, i)$ , where  $w \in A^*$  and  $i \in \text{Pos}(w)$ . Let  $\eta : A^* \rightarrow N$  be a morphism into a finite monoid and let  $k \in \mathbb{N}$ . Given  $w, w' \in A^*$ ,  $i \in \text{Pos}(w)$  and  $i' \in \text{Pos}(w')$ , we write,  $w, i \cong_{\eta, k} w', i'$  when:

$$\text{For every TL}[\eta] \text{ formula } \varphi \text{ with rank at most } k, \quad w, i \models \varphi \iff w', i' \models \varphi.$$

Clearly, the relations  $\cong_{\eta, k}$  are equivalences. Moreover, it is immediate from the definition and Lemma 9.4, that they have finite index. Finally, we also introduce equivalences which compare single words in  $A^*$ . Abusing terminology, we also write them  $\cong_{\eta, k}$ . Given  $w, w' \in A^*$ , we write  $w \cong_{\eta, k} w'$  if  $w, 0 \cong_{\eta, k} w', 0$ . Clearly, the relation  $\cong_{\eta, k}$  is an equivalence on  $A^*$ .

We complete this definition with a useful lemma. It presents an alternative definition of the equivalences  $\cong_{\eta, k}$ , which is convenient in proof arguments, using induction on  $k$ .

**LEMMA 9.5.** Let  $\eta : A^* \rightarrow N$  be a morphism into a finite monoid,  $k \in \mathbb{N}$ ,  $w, w' \in A^*$ ,  $i \in \text{Pos}(w)$  and  $i' \in \text{Pos}(w')$ . Then,  $w, i \cong_{\eta, k} w', i'$  if and only the five following conditions hold:

- We have  $w[i] = w'[i']$ .
- If  $k \geq 1$ , then for every  $j \in \text{Pos}(w)$  such that  $i < j$ , there exists  $j' \in \text{Pos}(w')$  such that  $i' < j'$ ,  $\eta(w(i, j)) = \eta(w'(i', j'))$  and  $w, j \cong_{\eta, k-1} w', j'$ .
- If  $k \geq 1$ , then for every  $j' \in \text{Pos}(w')$  such that  $i' < j'$ , there exists  $j \in \text{Pos}(w)$  such that  $i < j$ ,  $\eta(w(i, j)) = \eta(w'(i', j'))$  and  $w, j \cong_{\eta, k-1} w', j'$ .
- If  $k \geq 1$ , then for every  $j \in \text{Pos}(w)$  such that  $j < i$ , there exists  $j' \in \text{Pos}(w')$  such that  $j' < i'$ ,  $\eta(w(j, i)) = \eta(w'(j', i'))$  and  $w, j \cong_{\eta, k-1} w', j'$ .
- If  $k \geq 1$ , then for every  $j' \in \text{Pos}(w')$  such that  $j' < i'$ , there exists  $j \in \text{Pos}(w)$  such that  $j < i$ ,  $\eta(w(j, i)) = \eta(w'(j', i'))$  and  $w, j \cong_{\eta, k-1} w', j'$ .

**PROOF.** We start with the “only if” implication. Assume that  $w, i \cong_{\eta, k} w', i'$ . We show that the five conditions in the lemma hold. For the first one, we know that for every  $\ell \in A \cup \{\min, \max\}$ ,

“ $\ell$ ” is an atomic  $\text{TL}[\eta]$  formula, hence of rank 0. Therefore, our hypothesis implies that  $w, i \models \ell \Leftrightarrow w, i' \models \ell$  for every  $\ell \in A \cup \{\min, \max\}$ . This exactly says that  $w[i] = w'[i']$  and the first condition is proved. We turn to the four remaining conditions. By symmetry, we only detail one of them: we consider the second condition in the lemma. Hence we assume that  $k \geq 1$  and consider  $j \in \text{Pos}(w)$  such that  $i < j$ . We have to exhibit  $j' \in \text{Pos}(w')$  such that  $i' < j'$ ,  $\eta(w(i, j)) = \eta(w'(i', j'))$  and  $w, j \cong_{\eta, k-1} w', j'$ . We use  $w$  and  $j$  to build a  $\text{TL}[\eta]$  formula. Lemma 9.4 yields a finite set  $S$  of  $\text{TL}[\eta]$  formulas of rank at most  $k - 1$  such that every  $\text{TL}[\eta]$  formula  $\psi$  of rank at most  $k - 1$  is equivalent to some formula in  $S$ . Consider the set  $T = \{\psi \in S \mid w, j \models \psi\}$ . We define,

$$\varphi = \bigwedge_{\psi \in T} \psi.$$

Moreover, let  $s = \eta(w(i, j)) \in N$  (recall that  $i < j$ ) and let  $L = \eta^{-1}(s)$ . Clearly,  $w, j \models \varphi$  and it follows that  $w, i \models F_L \varphi$ . Moreover,  $\varphi$  has rank at most  $k - 1$  by definition, which means that  $F_L \varphi$  has rank at most  $k$ . Since  $w, i \cong_{\eta, k} w', i'$ , it follows that  $w', i' \models F_L \varphi$ . This yields  $j' \in \text{Pos}(w')$  such that  $i' < j'$ ,  $w'(i', j') \in L$ , i.e.,  $\eta(w'(i', j')) = \eta(w(i, j))$  and  $w', j' \models \varphi$ . It remains to show that  $w, j \cong_{\eta, k-1} w', j'$ . Let  $\psi$  be an  $\eta$ -formula of rank at most  $k - 1$ . We have to show that  $w, j \models \psi \Leftrightarrow w', j' \models \psi$ . For the left to right implication, if  $w, j \models \psi$ , then we know that  $T$  contains a formula equivalent to  $\psi$ . Recall that  $w', j' \models \varphi$ . By definition of  $\varphi$ , this yields  $w', j' \models \psi$ . For the converse implication, we prove the contrapositive. Assume that  $w, j \not\models \psi$ , i.e.,  $w, j \models \neg\psi$ . Since  $\neg\psi$  has rank at most  $k - 1$ , we know that  $T$  contains a formula equivalent to  $\neg\psi$ . Again by choice of  $\varphi$ , since  $w', j' \models \varphi$ , we also have  $w', j' \models \neg\psi$ , which implies that  $w', j' \not\models \psi$ , completing the proof.

Conversely, assume that the five conditions in the lemma are satisfied. We prove that  $w, i \cong_{\eta, k} w', i'$ . Given an arbitrary  $\text{TL}[\eta]$  formula  $\varphi$  of rank at most  $k$ , we have to prove that  $w, i \models \varphi \Leftrightarrow w', i' \models \varphi$ . We proceed by structural induction on  $\varphi$ . First, if  $\varphi$  is an atomic formula then either  $\varphi = \top$ ,  $\varphi = \perp$  or  $\varphi = \ell$  for some  $\ell \in A \cup \{\min, \max\}$ . In the first two cases, the result is trivial. In the last one, the first condition in the lemma states that  $w[i] = w'[i']$ . Hence, we have  $w, i \models \ell \Leftrightarrow w', i' \models \ell$  for every  $a \in A \cup \{\min, \max\}$ . We now consider Boolean connectives. If  $\varphi = \psi_1 \vee \psi_2$ , it follows from structural induction on  $\varphi$  that  $w, i \models \psi_1 \Leftrightarrow w', i' \models \psi_1$  and  $w, i \models \psi_2 \Leftrightarrow w', i' \models \psi_2$ . It is then immediate that  $w, i \models \varphi \Leftrightarrow w', i' \models \varphi$ . Similarly, if  $\varphi = \neg\psi$ , it follows from structural induction on  $\varphi$  that we have  $w, i \models \psi \Leftrightarrow w', i' \models \psi$ , whence immediately, that  $w, i \models \varphi \Leftrightarrow w', i' \models \varphi$ . It remains to treat the temporal operators, i.e., the cases where  $\varphi = F_L \psi$  or  $\varphi = P_L \psi$  where  $L \subseteq A^*$  is recognized by  $\eta$  (recall that  $\varphi$  is a  $\text{TL}[\eta]$  formula). Note that since  $\varphi$  has rank at most  $k$ , these cases may only happen when  $k \geq 1$  and  $\psi$  has rank at most  $k - 1$ . By symmetry, we only detail the case where  $\varphi = F_L \psi$ . Moreover, we only prove the implication  $w, i \models \varphi \Rightarrow w', i' \models \varphi$ , as the other one is proved symmetrically. Hence, we assume that  $w, i \models \varphi$ . Since  $\varphi = F_L \psi$ , it follows that there exists a position  $j$  in  $w$  such that  $i < j$ ,  $w(i, j) \in L$  and  $w, j \models \psi$ . The second condition in the lemma yields a position  $j'$  of  $w'$  such that

$i' < j'$ ,  $\eta(w(i, j)) = \eta(w'(i', j'))$  and  $w, j \cong_{\eta, k-1} w', j'$ . Since  $L$  is recognized by  $\eta$  and  $w(i, j) \in L$ , the equality  $\eta(w(i, j)) = \eta(w'(i', j'))$  implies that  $w'(i', j') \in L$ . Moreover, since  $\psi$  has rank at most  $k-1$  and both  $w, j \models \psi$  and  $w, j \cong_{\eta, k-1} w', j'$  hold, we obtain  $w', j' \models \psi$ . Altogether, since  $i' < j'$ , it follows that  $w', i' \models_{F_L} \psi$ , i.e., that  $w', i' \models \varphi$ , as desired. This concludes the proof. ■

It can be verified from Lemma 9.5 and a straightforward induction that the equivalences  $\cong_{\eta, k}$  on  $A^*$  are congruences. The detailed proof is left to the reader.

**LEMMA 9.6.** Consider a morphism  $\eta : A^* \rightarrow N$  into a finite monoid and  $k \in \mathbb{N}$ . For every  $u, v, u', v' \in A^*$  such that  $u \cong_{\eta, k} u'$  and  $v \cong_{\eta, k} v'$ , we have  $uv \cong_{\eta, k} u'v'$ .

We are ready to present the property of the equivalences  $\cong_{\eta, k}$  that we shall need in Section 10 to establish the correspondence with unambiguous polynomial closure.

**PROPOSITION 9.7.** Consider a morphism  $\eta : A^* \rightarrow N$  into a finite monoid, let  $e \in E(N)$  be an idempotent and let  $u, v, z \in \eta^{-1}(e)$ . For every  $k \in \mathbb{N}$ , the following property holds:

$$(z^k u z^{2k} v z^k)^k (z^k u z^{2k} v z^k)^k \cong_{\eta, k} (z^k u z^{2k} v z^k)^k z^k v z^k (z^k u z^{2k} v z^k)^k.$$

**PROOF.** We first show that we may restrict ourselves to the special case where  $u, v$  and  $z$  are single letters. To this aim, we consider a second independent alphabet  $B = \{a, b, c\}$  and we define  $\alpha : B^* \rightarrow A^*$  as the morphism given by  $\alpha(a) = u$ ,  $\alpha(b) = v$  and  $\alpha(c) = z$ . Finally, we let  $\delta = \eta \circ \alpha : B^* \rightarrow N$ . We have the following fact.

**FACT 9.8.** For all  $w, w' \in B^*$  and all  $k \in \mathbb{N}$ , if  $w \cong_{\delta, k} w'$ , then  $\alpha(w) \cong_{\eta, k} \alpha(w')$ .

**PROOF.** Let  $n = \max(\{|u|, |v|, |z|\})$ . For every  $x \in B^*$  and every  $h \leq n$ , we define a *partial* map  $f_{x, h} : \text{Pos}(x) \rightarrow \text{Pos}(\alpha(x))$ . We fix  $i \in \text{Pos}(x)$  for the definition. If  $i = 0$ , then  $f_{x, h}(0) = 0$  and if  $i = |x| + 1$ , then  $f_{x, h}(|x| + 1) = |\alpha(x)| + 1$ . Assume now that  $1 \leq i \leq |x|$  and let  $d = x[i] \in B$ . If  $h > |\alpha(d)|$ , then  $f_{x, h}(i)$  is undefined. Otherwise  $h \leq |\alpha(d)|$  and position  $i$  of  $x$  corresponds to a factor  $\alpha(d)$  in  $\alpha(x)$  which is made of more than  $h$  positions. We let  $f_{x, h}(i)$  be  $h$ -th position of this factor. Formally, this boils down to defining  $f_{x, h}(i) = |\alpha(x(0, i))| + h$ .

One may verify using Lemma 9.5 and induction on  $k \in \mathbb{N}$  that for every  $w, w' \in B^*$ , every  $i \in \text{Pos}(w)$  and every  $i' \in \text{Pos}(w')$ , if  $w, i \cong_{\delta, k} w', i'$ , then for all  $h \leq n$ ,  $f_{w, h}(i)$  is defined if and only if  $f_{w', h}(i')$  is defined and, in this case,  $\alpha(w), f_{w, h}(i) \cong_{\eta, k} \alpha(w'), f_{w', h}(i')$ . In particular, it follows that if  $w \cong_{\eta, k} w'$  (i.e.,  $w, 0 \cong_{\eta, k} w', 0$ ), then  $\alpha(w) \cong_{\eta, k} \alpha(w')$  as desired. ■

Note that by hypothesis on  $u, v$  and  $z$ , we have  $\delta(a) = \delta(b) = \delta(c) = e \in E(N)$ , which means that for every nonempty word  $w \in B^+$ , we have  $\delta(w) = e$  (on the other hand,  $\delta(\varepsilon) = 1_N$ , which might be distinct from  $e$ ). We prove that for every  $k \in \mathbb{N}$ , we have,

$$(c^k a c^{2k} b c^k)^k (c^k a c^{2k} b c^k)^k \cong_{\delta, k} (c^k a c^{2k} b c^k)^k c^k b c^k (c^k a c^{2k} b c^k)^k. \quad (12)$$

By Fact 9.8 and by definition of  $\alpha$ , this will yield the desired property of Proposition 9.7.

To prove (12), we establish by induction a more general property. We fix  $k$  for the proof, and we write  $x = c^k a c^k$  and  $y = c^k b c^k$ . Let  $n \in \mathbb{N}$  and consider a quadruple  $(w, i, w', i')$  where  $w, w' \in B^*$ ,  $i \in \text{Pos}(w)$  and  $i' \in \text{Pos}(w')$ . We say that  $(w, i, w', i')$  is an  $n$ -candidate if there exist  $h, \ell \geq n$  and  $w_1, w'_1 \in (x + y)^*$  such that  $w = (xy)^h w_1 (xy)^\ell$ ,  $w' = (xy)^h w'_1 (xy)^\ell$  and one of the three following properties is satisfied:

1. We have  $i \leq |(xy)^h|$ ,  $i' \leq |(xy)^h|$  and  $i = i'$ .
2. We have  $|(xy)^h w_1| < i$ ,  $|(xy)^h w'_1| < i'$  and  $i - |(xy)^h w_1| = i' - |(xy)^h w'_1|$ .
3. We have  $|(xy)^h| < i \leq |(xy)^h w_1|$ ,  $|(xy)^h| < i' \leq |(xy)^h w'_1|$ , and the following infixes are equal:  $w(i - n - 1, i + n + 1) = w'(i' - n - 1, i' + n + 1)$ .

We use Lemma 9.5 and induction on  $n$  to prove that for every  $n \leq k$  and every  $n$ -candidate  $(w, i, w', i')$ , we have  $w, i \cong_{\delta, n} w', i'$ . Since it is clear that  $((xy)^k (xy)^k, 0, (xy)^k y (xy)^k, 0)$  is a  $k$ -candidate (as Condition 1 in the definition holds), it will follow that the equivalence  $(xy)^k (xy)^k \cong_{\delta, k} (xy)^k y (xy)^k$  also holds. By definition of  $x$  and  $y$ , this is exactly (12).

Fix  $n \leq k$  and an  $n$ -candidate  $(w, i, w', i')$  for the proof. We have to show that  $w, i \cong_{\delta, n} w', i'$ . It is clear from the definition of an  $n$ -candidate that  $i$  and  $i'$  are either both unlabeled or share the same label in  $B$ . By Lemma 9.5, this concludes the proof when  $n = 0$ : we get  $w, i \cong_{\delta, 0} w', i'$ . We now assume that  $n \geq 1$ . In view of Lemma 9.5, there are four additional conditions to prove in this case. By symmetry, we only prove the first one and leave the others to the reader. Consider a position  $j$  of  $w$  such that  $i < j$ . We have to exhibit a position  $j'$  of  $w'$  such that  $i' < j'$ ,  $\eta(w(i, j)) = \eta(w'(i', j'))$  and  $w, j \cong_{\delta, n-1} w', j'$ . Observe that for the latter property (i.e.,  $w, j \cong_{\delta, n-1} w', j'$ ), it suffices to choose  $j'$  such that  $(w, j, w', j')$  is an  $(n - 1)$ -candidate: by induction on  $n$ , this implies that  $w, j \cong_{\eta, n-1} w', j'$ . Since  $(w, i, w', i')$  is an  $n$ -candidate, there exist  $h, \ell \geq n$  and  $w_1, w'_1 \in (x + y)^*$  such that  $w = (xy)^h w_1 (xy)^\ell$ ,  $w' = (xy)^h w'_1 (xy)^\ell$  and one of the three conditions in the definition holds. We consider several cases depending on the position  $j$  inside  $w$ .

Assume first that  $j \leq |(xy)^h|$ . Hence, we have  $i < j \leq |(xy)^h|$ . It follows that Condition 1 in the definition of the  $n$ -candidate  $(w, i, w', i')$  holds. Therefore, we have  $i' = i < |(xy)^h|$ . It now suffices to define  $j' = j$ . Clearly,  $i' < j'$  and  $w'(i', j') = w(i, j)$ . Moreover, one can check that  $(w, j, w', j')$  is an  $(n - 1)$ -candidate (in fact, it is even an  $n$ -candidate in this case) as Condition 1 in the definition holds.

We now assume that  $|(xy)^h| < j \leq |(xy)^h w_1 x y|$ . We consider two subcases depending on whether  $i + 1 = j$  or  $i + 1 < j$ . Assume first that  $i + 1 = j$ . In this case, we define  $j' = i' + 1$ . Clearly, we have  $i' < j'$  and  $w'(i', j') = w(i, j) = \varepsilon$ . Hence, we have to verify that  $(w, j, w', j')$  is an  $(n - 1)$ -candidate. We show that Condition 3 holds. By definition,  $h \geq n - 1$ ,  $\ell - 1 \geq n - 1$ ,  $w = (xy)^h w_1 x y (xy)^{\ell-1}$  and  $w' = (xy)^h w'_1 x y (xy)^{\ell-1}$ . Moreover,  $|(xy)^h| < j \leq |(xy)^h w_1 x y|$  and it can be verified from the definition of  $j'$  that  $|(xy)^h| < j' \leq |(xy)^h w'_1 x y|$ . Finally, one can also check that  $w(j - n, j + n) = w'(j - n, j' + n)$  from the hypothesis that  $(w, i, w', i')$  is an  $n$ -candidate,

since  $j = i + 1$  and  $j' = i' + 1$ . We conclude that  $(w, j, w', j')$  is an  $(n - 1)$ -candidate as Condition 3 holds. We turn to the second subcase: assume that  $i + 1 < j$ . By hypothesis,  $i' \leq |(xy)^h w'_1|$ . Hence, by definition of  $x, y$  and since  $n \leq k$ , one can verify that there exists a position  $j' \in \text{Pos}(w')$  such that  $|(xy)^h| < j' \leq |(xy)^h w'_1 x y|$ ,  $i' + 1 < j'$  and  $w(j - n, j + n) = w'(j - n, j' + n)$ . We have  $i' < j'$  by definition. Moreover, since  $i + 1 < j$  and  $i' + 1 < j'$ , we know that  $w'(i', j')$  and  $w(i, j)$  are nonempty, which yields  $\delta(w'(i', j')) = \delta(w(i, j)) = e$  by definition of  $\delta$ . Finally, it is immediate that  $(w, j, w', j')$  is an  $(n - 1)$ -candidate since Condition 3 again holds.

Finally, assume that  $|(xy)^h w_1 x y| < j$ . In this case, let  $j'$  be the unique position of  $w'$  such that  $|(xy)^h w'_1 x y| < j'$  and  $j - |(xy)^h w_1 x y| = j' - |(xy)^h w'_1 x y|$ . Since  $i < j$  and  $(w, i, w', i')$  is an  $n$ -candidate, one can check that  $i' < j'$  and that either  $i + 1 = j$  and  $i' + 1 = j'$  or  $i + 1 < j$  and  $i' + 1 < j'$ , which implies that  $\delta(w'(i', j')) = \delta(w(i, j))$ . Finally, it is straightforward to verify that  $(w, j, w', j')$  is a  $(n - 1)$ -candidate (in fact, it is even an  $n$ -candidate in this case) as Condition 2 in the definition holds, completing the proof. ■

## 9.2 Connection with two-variable first-order logic

We now prove the generic correspondence existing between two-variable first-order logic and unary temporal logic: we prove that  $\text{FO}^2(\mathbb{I}_C) = \text{TL}(C)$  for every Boolean algebra  $C$ . As we announced, this generalizes a theorem by Etessami, Vardi and Wilke [9], which states that  $\text{FO}^2(<) = \text{F} + \text{P}$  (which is the particular case  $C = \text{ST}$ ) and  $\text{FO}^2(<, +1) = \text{F} + \text{X} + \text{P} + \text{Y}$  (which is the particular case  $C = \text{ST}^+$ ).

**THEOREM 9.9.** *For every Boolean algebra  $C$ , we have  $\text{FO}^2(\mathbb{I}_C) = \text{TL}(C)$ .*

**PROOF.** We first prove the inclusion  $\text{TL}(C) \subseteq \text{FO}^2(\mathbb{I}_C)$ . We have to prove that for every  $\text{TL}[C]$  formula, there exists a sentence of  $\text{FO}^2(\mathbb{I}_C)$  defining the same language. Consider a  $\text{TL}[C]$  formula  $\varphi$ . We use structural induction to construct a formula  $\langle \varphi \rangle(x)$  of  $\text{FO}^2(\mathbb{I}_C)$  with at most one free variable “ $x$ ”, which satisfies the following property:

$$\text{Given } w \in A^* \text{ and a position } i \in \text{Pos}(w), \quad w \models \langle \varphi \rangle(i) \iff w, i \models \varphi. \quad (13)$$

Clearly, the language defined by  $\varphi$  is also defined by the sentence “ $\langle \varphi \rangle(\text{min})$ ” of  $\text{FO}^2(\mathbb{I}_C)$ . Hence, this proves  $\text{TL}(C) \subseteq \text{FO}^2(\mathbb{I}_C)$ , as announced.

It remains to construct  $\langle \varphi \rangle(x)$  from  $\varphi$ , which we do using structural induction on the  $\text{TL}[C]$  formula  $\varphi$ . Note that we only present the construction. That it satisfies (13) is straightforward to verify and left to the reader. First, if  $\varphi := \top$  or  $\varphi := \perp$ , then we define  $\langle \varphi \rangle(x) := \top$  and  $\langle \varphi \rangle(x) := \perp$  respectively. Moreover, if  $\varphi := \text{min}$  or  $\varphi := \text{max}$ , then we define  $\langle \varphi \rangle(x) := (x = \text{min})$  and  $\langle \varphi \rangle(x) := (x = \text{max})$  respectively. If  $\varphi := a$  for some letter  $a \in A$ , then we define  $\langle \varphi \rangle(x) := a(x)$ . Assume now that  $\varphi := F_L \psi$  for some  $L \in C$  and a smaller  $\text{TL}[C]$  formula  $\psi$ . In this case, we define  $\langle \varphi \rangle(x) := \exists y I_L(x, y) \wedge \langle \psi \rangle(y)$ . Finally, when  $\varphi := P_L \psi$  for some  $L \in C$  and a smaller

TL[C] formula  $\psi$ , we define  $\langle\psi\rangle(x) := \exists y I_L(y, x) \wedge \langle\psi\rangle(y)$ . This concludes the proof of the right to left inclusion in Theorem 9.9.

We turn to the inclusion  $\text{FO}^2(\mathbb{I}_C) \subseteq \text{TL}(C)$ . Let  $\varphi(x)$  be an  $\text{FO}^2(\mathbb{I}_C)$  formula with at most one free variable  $x$ . We prove that there exists a TL[C] formula  $[\varphi]$  satisfying the following property:

$$\text{Given } w \in A^* \text{ and a position } i \in \text{Pos}(w), \quad w, i \models [\varphi] \iff w \models \varphi(i). \quad (14)$$

Before we present the construction, let us explain why this implies  $\text{FO}^2(\mathbb{I}_C) \subseteq \text{TL}(C)$ . Let  $L \in \text{FO}^2(\mathbb{I}_C)$ . By definition,  $L$  is defined by a sentence  $\varphi$  of  $\text{FO}^2(\mathbb{I}_C)$ . Since  $\varphi$  has no free variables, the TL[C] formula  $[\varphi]$  given by (14) satisfies  $w, i \models [\varphi] \iff w \models \varphi$  for every  $w \in A^*$  and every  $i \in \text{Pos}(w)$ . In particular,  $[\varphi]$  defines  $L$  and we get  $L \in \text{TL}(C)$ .

It remains to prove that for every formula  $\varphi(x)$  of  $\text{FO}^2(\mathbb{I}_C)$  with at most one free variable  $x$ , there exists a TL[C] formula  $[\varphi]$  satisfying (14). We proceed by induction on the size of  $\varphi(x)$ . We start with the base case:  $\varphi(x)$  is an atomic formula. There are several cases. First, if  $\varphi(x) := \top$  or  $\varphi(x) := \perp$ , then we define  $[\varphi] := \top$  and  $[\varphi] := \perp$  respectively. If  $\varphi := a(x)$  for some letter  $a \in A$ , then we define  $[\varphi] := a$ . If  $\varphi := a(\min)$  or  $\varphi := a(\max)$  for some letter  $a \in A$ , then we define  $[\varphi] := \perp$ . We now consider the atomic formulas involving equality. Since there is only one free variable  $x$  and equality is commutative, this boils down to four cases. If  $\varphi := (x = x)$ , we define  $[\varphi] := \top$ . If  $\varphi := (x = \min)$ , we define  $[\varphi] := \min$ . If  $\varphi := (x = \max)$ , we define  $[\varphi] := \max$ . Finally, if  $\varphi := (\min = \max)$ , we define  $[\varphi] := \perp$ . It now remains to consider the atomic formulas involving a binary predicate  $I_L$  for some  $L \in C$ . Since  $x$  is the only variable which is (possibly) free in  $\varphi(x)$ , we have to consider the following cases. If  $\varphi$  is either  $I_L(x, x)$ ,  $I_L(\max, x)$ ,  $I_L(x, \min)$  or  $I_L(\max, \min)$ , we simply define  $[\varphi] := \perp$ . If we have  $\varphi := I_L(\min, x)$ , we define  $[\varphi] := P_L \min$ . If we have  $\varphi := I_L(x, \max)$ , we define  $[\varphi] := F_L \max$ . Finally, when  $\varphi := I_L(\min, \max)$ , we define  $[\varphi] := (\min \wedge F_L \max) \vee P_{A^*} (\min \wedge F_L \max)$  (note that  $A^* \in C$  since  $C$  is a Boolean algebra). This concludes the case of atomic formulas. The Boolean connectives are handled in the obvious way: we let  $[\psi_1 \vee \psi_2] = [\psi_1] \vee [\psi_2]$ ,  $[\psi_1 \wedge \psi_2] = [\psi_1] \wedge [\psi_2]$  and  $[\neg\psi] = \neg[\psi]$ .

It remains to treat quantification over a second variable  $y$ , i.e.,  $\varphi(x) := \exists y \gamma(x, y)$  (since  $\forall$  has the same semantics as  $\neg\exists\neg$ , we may assume without loss of generality that all quantifiers in  $\varphi$  are existential). We cannot apply induction to  $\gamma$  since it has *two* free variables. To solve this problem, we first need to put  $\gamma$  into normal form. First, we define  $\mathbf{L}$  as the finite set of all languages  $L \in C$  such that  $\gamma$  contains an atomic formula  $I_L(x, y)$  or  $I_L(y, x)$  inside which the occurrences of  $x$  and  $y$  are free. Using  $\mathbf{L}$ , we define an equivalence  $\sim$  over  $A^*$ : for  $u, v \in A^*$ , we write  $u \sim v$  when  $u \in L \iff v \in L$  for every  $L \in \mathbf{L}$ . Since  $\mathbf{L}$  is finite,  $\sim$  has finite index. Moreover, since every  $L \in \mathbf{L}$  belongs to  $C$ , which is a Boolean algebra by hypothesis, the  $\sim$ -classes belong to  $C$  as well. We say that a formula is an *L-profile* when it is of the form  $I_H(x, y)$ ,  $I_H(y, x)$  or  $x = y$  where  $H \in C$  is a  $\sim$ -class. Note that since  $\sim$  has finite index, there are finitely many L-profiles.

Finally, we say that an  $\text{FO}^2(\mathbb{I}_C)$  formula with at most two free-variables  $x$  and  $y$  is **L-normalized** if it has the following form:

$$\bigwedge_{1 \leq i \leq n} (\zeta_i(x) \Leftrightarrow t_i) \wedge \pi(x, y) \wedge \psi(y),$$

where  $n \in \mathbb{N}$ ,  $\zeta_1(x), \dots, \zeta_n(x)$  are  $\text{FO}^2(\mathbb{I}_C)$  formulas *smaller than*  $\gamma$  whose only free variable is  $x$ ,  $t_1, \dots, t_n \in \{\top, \perp\}$ ,  $\pi(x, y)$  is an **L-profile**, and  $\psi(y)$  is an  $\text{FO}^2(\mathbb{I}_C)$  formula *smaller than*  $\gamma$  whose only free variable is  $y$ . The argument is based on the following fact.

**FACT 9.10.** The formula  $\gamma(x, y)$  is equivalent to a finite disjunction of **L-normalized** formulas.

Before we prove Fact 9.10, let us use it to complete the construction of  $[\varphi]$ . Recall that  $\varphi(x) := \exists y \gamma(x, y)$ . Since disjunctions and existential quantifications commute, it follows from Fact 9.10 that  $\varphi(x)$  is equivalent to a finite disjunction of the form

$$\bigvee_{1 \leq j \leq m} \varphi_j(x).$$

where each formula  $\varphi_j(x)$  is of form  $\exists y \xi_j(x, y)$  with  $\xi_j(x, y)$  an **L-normalized** formula. Therefore, we concentrate on the formulas  $\varphi_j$ : if we construct  $\text{TL}[C]$  formulas  $[\varphi_1], \dots, [\varphi_m]$  satisfying (14) for  $\varphi_1, \dots, \varphi_m$ , it will then suffice to define:

$$[\varphi] = \bigvee_{1 \leq j \leq m} [\varphi_j].$$

Consider one of the formulas  $\varphi_j$ . We use induction to build  $[\varphi_j]$ . By definition of normalized formulas,  $\varphi_j(x)$  has the following form:

$$\varphi_j(x) = \exists y \bigwedge_{1 \leq i \leq n} (\zeta_i(x) \Leftrightarrow t_i) \wedge \pi(x, y) \wedge \psi(y).$$

with  $n \in \mathbb{N}$ ,  $\zeta_1(x), \dots, \zeta_n(x)$  are  $\text{FO}^2(\mathbb{I}_C)$  formulas *smaller than*  $\gamma$  whose only free variable is  $x$ ,  $t_1, \dots, t_n \in \{\top, \perp\}$ ,  $\pi(x, y)$  an **L-profile** and  $\psi(y)$  is an  $\text{FO}^2(\mathbb{I}_C)$  formula *smaller than*  $\gamma$  whose only free variable is  $y$ . We may therefore apply induction to the subformulas  $\zeta_1(x), \dots, \zeta_n(x)$  and  $\psi(y)$ , which yields  $\text{FO}^2(\mathbb{I}_C)$  formulas  $[\zeta_1], \dots, [\zeta_n]$  and  $[\psi]$  satisfying (14). We consider three cases depending on the **L-profile**  $\pi(x, y)$ . First, if  $\pi(x, y) := "I_H(x, y)"$  for some  $\sim$ -class  $H \in C$ , we define,

$$[\varphi_j] = \bigwedge_{1 \leq i \leq n} ([\zeta_i] \Leftrightarrow t_i) \wedge F_H [\psi].$$

Second, if  $\pi(x, y) := "I_H(y, x)"$  for some  $\sim$ -class  $H \in C$ , we define,

$$[\varphi_j] = \bigwedge_{1 \leq i \leq n} ([\zeta_i] \Leftrightarrow t_i) \wedge P_H [\psi].$$

Finally, if  $\pi(x, y) := "x = y"$  for some  $\sim$ -class  $H \in \mathcal{C}$ , we define,

$$[\varphi_j] = \bigwedge_{1 \leq i \leq n} ([\zeta_i] \Leftrightarrow t_i) \wedge [\psi].$$

One may verify that this definition satisfies (14), which concludes the proof for the construction of  $[\varphi]$ .

It remains to prove Fact 9.10. By definition,  $\gamma(x, y)$  is a Boolean combination of atomic formulas and existential quantifications, *i.e.*, formulas of the form  $\exists z \psi(x, y, z)$ . Note that since  $\gamma(x, y)$  is  $\text{FO}^2$  and already contains the two variables  $x$  and  $y$ , we know that in the latter case, the quantified variable  $z$  is either  $x$  or  $y$ , which means that there remains in the formula a *single* free variable (namely,  $x$  if  $z = y$  and  $y$  if  $z = x$ ). The important point resulting from this discussion is that the only subformulas in the Boolean combination involving *both*  $x$  and  $y$  as free variables are *atomic*. In summary, we conclude that  $\gamma(x, y)$  is a Boolean combination of three kinds of subformulas:

1. Atomic formulas involving *both*  $x$  and  $y$ : by definition, they are of the form  $I_L(x, y)$  or  $I_L(y, x)$  for  $L \in \mathbf{L}$ .
2. Formulas  $\zeta_1(x), \dots, \zeta_n(x)$  whose only free variable is  $x$ .
3. Formulas  $\chi_1(y), \dots, \chi_m(y)$  whose only free variable is  $y$ .

Intuitively, the main idea is now to get rid of the first two kinds of subformulas from  $\gamma(x, y)$  by making a disjunction over all possible assignments of truth values for  $\zeta_1(x), \dots, \zeta_n(x)$  and all possible formulas  $I_L(x, y)$  or  $I_L(y, x)$  for  $L \in \mathbf{L}$ .

Let  $T = \{\perp, \top\}^n$ . Intuitively, we view each tuple  $\tau = (t_1, \dots, t_n) \in T$  as an assignment of truth values for the subformulas  $\zeta_1(x), \dots, \zeta_n(x)$ . Moreover, let  $P$  be the set of all  $\mathbf{L}$ -profiles. By definition, an  $\mathbf{L}$ -profile  $\pi(x, y)$  determines the truth value of every atomic formula  $I_L(x, y)$  or  $I_L(y, x)$  for  $L \in \mathbf{L}$ . Indeed, when  $\pi(x, y) := "x = y"$  is satisfied, then all atomic formulas  $I_L(x, y)$  or  $I_L(y, x)$  are false. When  $\pi(x, y) := "I_H(x, y)"$  is satisfied for some  $\sim$ -class  $H$ , then all formulas  $I_L(y, x)$  are false and given  $L \in \mathbf{L}$ ,  $I_L(x, y)$  is true if and only if  $H \subseteq L$  (by definition of the equivalence  $\sim$  from  $\mathbf{L}$ ). Symmetrically, when  $\pi(x, y) := "I_H(y, x)"$  is satisfied for some  $\sim$ -class  $H$ , then all formulas  $I_L(x, y)$  are false and given  $L \in \mathbf{L}$ ,  $I_L(y, x)$  is true if and only if  $H \subseteq L$ . For every  $\tau = (t_1, \dots, t_n) \in T$  and  $\pi \in P$ , we denote by  $\pi(\tau(\gamma(x, y)))$  the formula obtained from  $\gamma(x, y)$  by replacing each subformula  $\zeta_i(x)$  in the Boolean combination by the truth value  $t_i \in \{\perp, \top\}$  and each atomic formula  $I_L(x, y)$  or  $I_L(y, x)$  for  $L \in \mathbf{L}$  by its truth value as given by  $\pi$ . Observe that the only free variable in  $\pi(\tau(\gamma(x, y)))$  is  $y$  since we replaced all subformulas involving  $x$  in the Boolean combination by truth values. One may now verify that  $\gamma(x, y)$  is equivalent to the following formula:

$$\bigvee_{\tau=(t_1, \dots, t_n) \in T} \bigvee_{\pi \in P} \left( \bigwedge_{i \leq n} (\zeta_i(x) \Leftrightarrow t_i) \wedge \pi(x, y) \wedge \pi(\tau(\gamma(x, y))) \right).$$

By definition, this is a disjunction of normalized formulas. Note also that each formula  $\zeta_i(x)$  is smaller than  $\gamma(x, y)$ , as well as  $\psi(y) = \pi(\tau(\gamma(x, y)))$ . This concludes the proof of Fact 9.10. ■

### 9.3 Connection with unambiguous polynomial closure

We now prove that if  $C$  is a prevariety, the inclusion  $UPol(BPol(C)) \subseteq TL(C)$  holds. Let us point out that this inclusion is *strict* in general. Yet, we shall prove in the next section that when  $C$  is either a prevariety of group languages  $\mathcal{G}$  or its well-suited extension  $\mathcal{G}^+$ , the inclusion is in fact an equality: in that case, we have  $UPol(BPol(C)) = TL(C)$ .

**REMARK 9.11.** The cases when  $C$  is either  $\mathcal{G}$  or  $\mathcal{G}^+$  for a prevariety of group languages  $\mathcal{G}$  cover most of the situations when  $TL(C)$  corresponds to a natural logical class. Indeed, as  $TL(C) = FO^2(\mathbb{I}_C)$  by Theorem 9.9, it follows from the results of Section 3 that when  $C$  is either  $\mathcal{G}$  or  $\mathcal{G}^+$ , we obtain the variant  $FO^2(<, \mathbb{P}_{\mathcal{G}})$  and  $FO^2(<, +1, \mathbb{P}_{\mathcal{G}})$  of  $FO^2$ . Yet, there is a special variant of  $FO^2$  that is not covered by these cases. It was considered by Krebs, Lodaya, Pandya and Straubing [12]. It is called “two-variable first-order logic with between predicates” and it is simple to verify that it corresponds to  $FO^2(\mathbb{I}_{AT})$ . It is proved in [12] that this variant has decidable membership but this is based on a specialized proof which is independent of the techniques that we use here.

We may now prove the generic inclusion  $UPol(BPol(C)) \subseteq TL(C)$ . The proof argument is based on Theorem 4.29 which yields  $UPol(BPol(C)) = WAPol(BPol(C))$ . With this equality in hand, it suffices to prove that  $WAPol(BPol(C)) \subseteq TL(C)$ .

**PROPOSITION 9.12.** Let  $C$  be a prevariety. Then,  $UPol(BPol(C)) \subseteq TL(C)$ .

**PROOF.** By Corollary 4.6,  $BPol(C)$  is a prevariety. Hence,  $UPol(BPol(C)) = WAPol(BPol(C))$  by Theorem 4.29 and it suffices to prove that  $WAPol(BPol(C)) \subseteq TL(C)$ . By definition of  $WAPol$ , this boils down to proving that  $TL(C)$  contains  $BPol(C)$  and is closed under disjoint union and left/right  $BPol(C)$ -deterministic marked concatenation. Closure under union is immediate by definition. For the other properties, we use the following fact.

**FACT 9.13.** Let  $K \in BPol(C)$ . There exist two  $TL[C]$  formulas  $\xi_K^l$  and  $\xi_K^r$  such that for every word  $w \in A^*$  and every position  $i \in Pos(w)$ , we have,

$$\begin{aligned} w, i \models \xi_K^l &\iff i \geq 1 \text{ and } w(0, i) \in K. \\ w, i \models \xi_K^r &\iff i \leq |w| \text{ and } w(i, |w| + 1) \in K. \end{aligned}$$

**PROOF.** We prove the existence of  $\xi_K^r$  (the argument for  $\xi_K^l$  is symmetrical and left to the reader). By definition,  $K \in BPol(C)$  is a Boolean combination of languages  $K_0 a_1 K_1 \cdots a_n K_n$  where  $a_1, \dots, a_n \in A$  and  $K_0, \dots, K_n \in C$ . Therefore, since we may use Boolean connectives

freely in  $\text{TL}[C]$  formulas, we may assume without loss of generality that  $K$  itself is of the form  $K_0 a_1 K_1 \cdots a_n K_n$ . In that case, it suffices to define,

$$\xi_K^r = F_{K_0} (a_1 \wedge F_{K_1} (a_2 \wedge F_{K_2} (\cdots a_n \wedge F_{K_n} \text{max}))).$$

By definition,  $\xi_K^r$  is a  $\text{TL}[C]$  formula. Moreover, one can verify that it satisfies the desired property, concluding the proof. ■

Clearly, Fact 9.13 implies that  $\text{TL}(C)$  contains  $\text{BPol}(C)$ : every language  $K \in \text{BPol}(C)$  is defined by the  $\text{TL}[C]$  formula  $\xi_K^r$ . It remains to prove that  $\text{TL}(C)$  is closed under left/right  $\text{BPol}(C)$ -deterministic marked concatenation. By symmetry, we concentrate on *left*  $\text{BPol}(C)$ -deterministic marked concatenation. Let  $H, L \in \text{TL}(C)$  and  $a \in A$  such that  $HaL$  is left  $\text{BPol}(C)$ -deterministic. We prove that  $HaL \in \text{TL}(C)$ . By definition, we have  $K \in \text{BPol}(C)$  such that  $H \subseteq K$  and  $KaL$  is left deterministic. We write  $\psi$  for the  $\text{TL}[C]$  formula  $a \wedge \xi_K^\ell$  (where  $\xi_K^\ell$  is given by Fact 9.13). Since  $KaL$  is left deterministic, it follows from Lemma 4.14 that  $KaA^*$  is unambiguous. Hence, by definition of  $\psi$ , for every word  $w \in A^*$ , there exists *at most one* position  $i \in \text{Pos}(w)$  such that  $w, i \models \psi$ . Since  $H \subseteq K$ , it follows that for every  $w \in A^*$ , we have  $w \in HaL$  if and only if  $w$  satisfies the three properties:

1. there exists  $i \in \text{Pos}(w)$  such that  $w, i \models \psi$  (the position  $i$  is unique as explained above).
2. The prefix  $w(0, i)$  belongs to  $H$ .
3. The suffix  $w(i, |w| + 1)$  belongs to  $L$ .

Therefore, it suffices to prove that these three properties may be expressed using a  $\text{TL}[C]$  formula. This is simple for the first property: it is expressed by the formula  $F \psi$ . It remains to prove that the other two properties can be defined as well. Let us start with the second one. Since  $H \in \text{TL}(C)$ , it is defined by a  $\text{TL}[C]$  formula  $\varphi_H$ . We modify it to construct another formula  $\varphi'_H$  expressing the second property. Given  $w \in A^*$ , we restrict the evaluation of  $\varphi_H$  to the positions  $i \in \text{Pos}(w)$  such that  $i$  is either strictly smaller than the unique one satisfying  $\psi$  or  $i = |w| + 1$ . More precisely, we build  $\varphi'_H$  by applying the two following modifications to  $\varphi_H$ :

1. We recursively replace each subformula of the form  $P_V \zeta$  with  $V \in C$  by,

$$(\text{max} \wedge P (\psi \wedge P_V \zeta)) \vee (F \psi \wedge P_V \zeta).$$

2. We recursively replace each subformula of the form  $F_V \zeta$  with  $V \in C$  by,

$$(F_V (\zeta \wedge F \psi)) \vee (F_V (\psi \wedge F (\text{max} \wedge \zeta))).$$

It remains to handle the third property. Since  $L \in \text{TL}(C)$ , there exists a  $\text{TL}[C]$  formula  $\varphi_L$  defining  $L$ . We construct a formula  $\varphi'_L$  expressing the third property above.

Given  $w \in A^*$ , we restrict the evaluation of  $\varphi_L$  to the positions  $i \in \text{Pos}(w)$  such that  $i$  is either strictly larger than the unique one satisfying  $\psi$  or  $i = 0$ . More precisely, we build  $\varphi'_L$  by applying the two following modifications to  $\varphi_L$ :

1. We recursively replace each subformula of the form  $P_V \zeta$  with  $V \in C$  by,

$$(P_V (\zeta \wedge P \psi)) \vee (P_V (\psi \wedge P (\min \wedge \zeta))).$$

2. We recursively replace each subformula of the form  $F_V \zeta$  with  $V \in C$  by,

$$(\min \wedge F (\psi \wedge F_V \zeta)) \vee (P \psi \wedge F_V \zeta).$$

The language  $HaL$  is now defined by the  $\text{TL}[C]$  formula  $(F \psi) \wedge \varphi'_H \wedge \varphi'_L$ . This concludes the proof of Proposition 9.12. ■

## 10. Logical characterizations of unambiguous polynomial closure

This final section details the logical characterizations of unambiguous polynomial closure. Let us first summarize what we already know. By Theorem 4.9, when  $C$  is a prevariety, we have  $\Delta_2(\mathbb{I}_C) = \text{Pol}(\text{BPol}(C)) \cap \text{coPol}(\text{BPol}(C))$ . Moreover, since  $\text{BPol}(C)$  is a prevariety by Corollary 4.6, we get from Theorem 4.21 that,

$$\Delta_2(\mathbb{I}_C) = \text{Pol}(\text{BPol}(C)) \cap \text{coPol}(\text{BPol}(C)) = \text{UPol}(\text{BPol}(C)).$$

Independently, we proved in Theorem 9.9 that the variant  $\text{FO}^2(\mathbb{I}_C)$  of two-variable first-order logic corresponds to the unary temporal logic  $\text{TL}(C)$  (this holds as soon as  $C$  is a Boolean algebra). Finally, we proved the inclusion  $\text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C)$  in Proposition 9.12 for every prevariety  $C$ . Altogether, it follows that when  $C$  is a prevariety, we have

$$\Delta_2(\mathbb{I}_C) = \text{Pol}(\text{BPol}(C)) \cap \text{coPol}(\text{BPol}(C)) = \text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C) = \text{FO}^2(\mathbb{I}_C).$$

In general, the inclusion is strict. Yet, we prove in this section that when  $C$  is either a prevariety of group languages  $\mathcal{G}$  or its well-suited extension  $\mathcal{G}^+$ , the converse inclusion holds. Since the presentation of logical classes can be simplified in these cases (see Lemma 3.3 and Lemma 9.2), we get the following generic results for every prevariety of group languages  $\mathcal{G}$ :

$$\begin{aligned} \text{UPol}(\text{BPol}(\mathcal{G})) &= \Delta_2(<, \mathbb{P}_{\mathcal{G}}) = \text{FO}^2(<, \mathbb{P}_{\mathcal{G}}) = \text{TL}(\mathcal{G}). \\ \text{UPol}(\text{BPol}(\mathcal{G}^+)) &= \Delta_2(<, +1, \mathbb{P}_{\mathcal{G}}) = \text{FO}^2(<, +1, \mathbb{P}_{\mathcal{G}}) = \text{TLX}(\mathcal{G}). \end{aligned}$$

The proofs of the missing inclusions are based on Theorem 6.7, the algebraic characterization of  $\text{Pol}(\text{BPol}(C)) \cap \text{coPol}(\text{BPol}(C)) = \Delta_2(\mathbb{I}_C)$ . We use Proposition 9.7 to prove that it is satisfied by the class  $\text{TL}(C)$  when  $C$  is a prevariety of group languages  $\mathcal{G}$  or its well-suited extension  $\mathcal{G}^+$ . Actually, we simplify the generic characterization presented in Theorem 6.7 and present

specialized characterizations for these two special cases. In particular, this yields generic characterizations of  $\text{FO}^2(<, \mathbb{P}_{\mathcal{G}})$  and  $\text{FO}^2(<, +1, \mathbb{P}_{\mathcal{G}})$ . They generalize well-known results for particular instances of these logical classes. Let us point out that here, we directly characterize the languages of these classes (using a property of their syntactic morphism) rather than characterizing the morphisms associated to the class. This is because in this case, we do need to use the characterizations as subresults.

## 10.1 Group languages

We first consider the classes  $\text{UPol}(\text{BPol}(\mathcal{G}))$  where  $\mathcal{G}$  is a prevariety of group languages. The algebraic characterization is based on the class of finite monoids  $\mathbf{DA}$ , whose connection with unambiguous polynomial closure [35] and two-variable first-order logic [39] is well known. We use a definition based on an equation (see [38] for details). A finite monoid  $N$  belongs to  $\mathbf{DA}$  when it satisfies the following equation:

$$(st)^\omega = (st)^\omega t (st)^\omega \quad \text{for every } s, t \in N. \quad (15)$$

We are now ready to present the generic characterization for classes of the form  $\text{UPol}(\text{BPol}(\mathcal{G}))$ .

**THEOREM 10.1.** *Let  $\mathcal{G}$  be a prevariety of group languages and let  $L$  be a regular language. The following properties are equivalent:*

1.  $L \in \text{UPol}(\text{BPol}(\mathcal{G}))$ .
2.  $L \in \text{Pol}(\text{BPol}(\mathcal{G})) \cap \text{coPol}(\text{BPol}(\mathcal{G}))$ .
3.  $L \in \Delta_2(<, \mathbb{P}_{\mathcal{G}})$ .
4.  $L \in \text{FO}^2(<, \mathbb{P}_{\mathcal{G}})$ .
5.  $L \in \text{TL}(\mathcal{G})$ .
6. *The  $\mathcal{G}$ -kernel of the syntactic morphism of  $L$  belongs to  $\mathbf{DA}$ .*

**PROOF.** By Lemma 3.3, we have  $\Delta_2(<, \mathbb{P}_{\mathcal{G}}) = \Delta_2(\mathbb{I}_{\mathcal{G}})$ . Therefore, (2)  $\Leftrightarrow$  (3) follows from Theorem 4.9. Moreover, (1)  $\Leftrightarrow$  (2) is given by Theorem 4.21. Since  $\text{FO}^2(<, \mathbb{P}_{\mathcal{G}}) = \text{FO}^2(\mathbb{I}_{\mathcal{G}})$  by Lemma 3.3, the equivalence (4)  $\Leftrightarrow$  (5) is immediate from Theorem 9.9. Finally, the implication (1)  $\Rightarrow$  (5) follows from Proposition 9.12. We now prove independently that (6)  $\Rightarrow$  (2) and (5)  $\Rightarrow$  (6), which completes the argument. Let  $\alpha : A^* \rightarrow M$  be the syntactic morphism of  $L$  and let  $N \subseteq M$  be its  $\mathcal{G}$ -kernel.

Let us start with the implication (6)  $\Rightarrow$  (2). We assume that  $N$  belongs to  $\mathbf{DA}$  and show that  $L \in \text{Pol}(\text{BPol}(\mathcal{G})) \cap \text{coPol}(\text{BPol}(\mathcal{G}))$ . Since  $L$  is recognized by its syntactic morphism  $\alpha$ , it suffices to show that  $\alpha$  is a  $(\text{Pol}(\text{BPol}(\mathcal{G})) \cap \text{coPol}(\text{BPol}(\mathcal{G})))$ -morphism. By Theorem 6.7, this boils down to proving that for every  $s, t \in M$  and  $e \in E(M)$  such that  $(e, s) \in M^2$  is a  $\mathcal{G}$ -pair, we have  $(eset)^{\omega+1} = (eset)^\omega et (eset)^\omega$ . By Lemma 5.13, we have  $es \in N$  and  $et(eset)^{2\omega-1} \in N$ .

Since  $N$  belongs to **DA**, it follows from (15) that,

$$(eset(eset)^{2\omega-1})^\omega = (eset(eset)^{2\omega-1})^\omega et(eset)^{2\omega-1}(eset(eset)^{2\omega-1})^\omega.$$

This exactly says that  $(eset)^\omega = (eset)^\omega et(eset)^{3\omega-1}$ . It now suffices to multiply by  $eset$  on the right to get  $(eset)^{\omega+1} = (eset)^\omega et(eset)^\omega$ , as desired.

Finally, we prove that (5)  $\Rightarrow$  (6). We assume that  $L \in \text{TL}(\mathcal{G})$  and show that  $N \in \mathbf{DA}$ , i.e., that  $N$  satisfies (15). Let  $s, t \in N$ . We prove that  $(st)^\omega = (st)^\omega t(st)^\omega$ . By hypothesis,  $L$  is defined by a formula  $\varphi_L$  of  $\text{TL}[\mathcal{G}]$ . Let  $k \in \mathbb{N}$  be the rank of  $\varphi_L$ . Fact 9.3 yields a  $\mathcal{G}$ -morphism  $\eta : A^* \rightarrow G$  such that  $\varphi_L$  is a  $\text{TL}[\eta]$  formula. Moreover, since  $s, t$  belong to the  $\mathcal{G}$ -kernel  $N$  of  $\alpha$  and  $\eta$  is a  $\mathcal{G}$ -morphism, Lemma 5.4 yields  $u, v \in A^*$  such that  $\eta(u) = \eta(v) = 1_G$ ,  $\alpha(u) = s$  and  $\alpha(v) = t$ . Hence, since  $\varphi_L$  is an  $\eta$ -formula of rank  $k$ , it follows from Lemma 9.6 and Proposition 9.7 (applied for  $z = \varepsilon$  which also maps to  $1_G$ ) that  $x(uv)^k(uv)^k y \in L \Leftrightarrow x(uv)^k v(uv)^k y \in L$  for all  $x, y \in A^*$ . In other words,  $(uv)^k(uv)^k$  and  $(uv)^k v(uv)^k$  are equivalent for the syntactic congruence of  $L$ . Since  $\alpha$  is the syntactic morphism of  $L$ , we get  $\alpha((uv)^k(uv)^k) = \alpha((uv)^k v(uv)^k)$ . By definition of  $u$  and  $v$ , it follows that  $(st)^k(st)^k = (st)^k t(st)^k$ . It now suffices to multiply by enough copies of  $st$  on the left and on the right to obtain  $(st)^\omega = (st)^\omega t(st)^\omega$ , as desired.  $\blacksquare$

Given as input a regular language  $L$ , one can compute its syntactic morphism. Moreover, by Lemma 5.3, the  $\mathcal{G}$ -kernel of this morphism can be computed when  $\mathcal{G}$ -separation is decidable. It is then simple to decide whether it belongs to **DA**: this boils down to checking whether (15) holds by testing all possible combinations for  $s$  and  $t$ . By Theorem 10.1, this decides if  $L \in \text{UPol}(\text{BPol}(\mathcal{G}))$ . Altogether, we obtain the following corollary.

**COROLLARY 10.2.** Let  $\mathcal{G}$  be a prevariety of group languages with decidable separation. Then, membership is decidable for the class  $\text{FO}^2(<, \mathbb{P}_{\mathcal{G}}) = \text{TL}(\mathcal{G}) = \Delta_2(<, \mathbb{P}_{\mathcal{G}}) = \text{UPol}(\text{BPol}(\mathcal{G}))$ .

Recall that separation is decidable for the four standard prevarieties of group languages **ST**, **MOD**, **AMT** and **GR** that we presented in Section 2.3. Therefore, Corollary 10.2 applies in these cases. In particular, we get the decidability of membership for the logical classes  $\text{FO}^2(<)$ ,  $\text{FO}^2(<, \text{MOD})$ ,  $\text{FO}^2(<, \text{AMOD})$  and  $\text{FO}^2(<, \mathbb{P}_{\text{GR}})$ .

Again, Theorem 10.1 generalizes classic results. The most prominent one is for  $\mathcal{G} = \text{ST}$ . Since  $\text{ST} = \{\emptyset, A^*\}$ , it is easy to check that the **ST**-kernel of a surjective morphism  $\alpha : A^* \rightarrow M$  is the whole monoid  $M$ . Also by Lemma 4.25, we have  $\text{UPol}(\text{BPol}(\text{ST})) = \text{UPol}(\text{AT})$ . Therefore, in this case, Theorem 10.1 yields that  $\text{UPol}(\text{AT}) = \Delta_2(<) = \text{FO}^2(<) = \text{F} + \text{P}$  and that a regular language belongs to this class if and only if its syntactic monoid belongs to **DA**. Historically, this was proved by combining several independent results. First, the correspondence between  $\text{UPol}(\text{AT})$  and **DA** is due to Schützenberger [35]. Then, it was shown by Pin and Weil [20] that  $\text{UPol}(\text{AT}) = \Delta_2(<)$ . As we explained in the previous section, that  $\text{FO}^2(<) = \text{F} + \text{P}$  was proved by Etessami, Vardi and Wilke [9]. Finally, the correspondence between  $\text{FO}^2(<)$  and **DA** is due to Thérien and Wilke [39].

Another well-known application of Theorem 10.1 is the case where  $\mathcal{G} = \text{MOD}$ . In particular, the theorem implies that  $\Delta_2(<, \text{MOD}) = \text{FO}^2(<, \text{MOD})$  and that a regular language belongs to this class if and only if the MOD-kernel of its syntactic morphism belongs to **DA**. The correspondence between  $\text{FO}^2(<, \text{MOD})$  and the membership of the MOD-kernel of the syntactic morphism to **DA** is due to Dartois and Paperman [7], while the equality  $\Delta_2(<, \text{MOD}) = \text{FO}^2(<, \text{MOD})$  is due to Kufleitner and Walter [13].

**REMARK 10.3.** While membership for  $\text{UPol}(\text{BPol}(\mathcal{G}))$  boils down to  $\mathcal{G}$ -separation, not much is known about *separation* and *covering* for  $\text{UPol}(\text{BPol}(\mathcal{G}))$ . The only class of this kind for which separation and covering are known to be decidable is  $\text{UPol}(\text{BPol}(\text{ST})) = \text{FO}^2(<)$ . This is the class  $\text{UPol}(\text{AT})$  by Lemma 4.25. Since AT is a *finite* class,  $\text{UPol}(\text{AT})$ -covering is decidable by Corollary 8.11.

## 10.2 Well-suited extensions

We now consider classes of the form  $\text{UPol}(\text{BPol}(\mathcal{G}^+))$ , where  $\mathcal{G}^+$  is the well-suited extension of an arbitrary prevariety of group languages  $\mathcal{G}$ . In this case, the characterization is based on the class of finite semigroups **LDA**. Again, we use a definition which is based on an equation. A finite semigroup  $S$  belongs to **LDA** when it satisfies the following equation:

$$(esete)^\omega = (esete)^\omega ete(esete)^\omega \quad \text{for every } s, t \in S \text{ and } e \in E(S). \quad (16)$$

We now present the generic statement characterization of  $\text{UPol}(\text{BPol}(\mathcal{G}^+))$ .

**THEOREM 10.4.** *Let  $\mathcal{G}$  be a prevariety of group languages and let  $L$  be a regular language. The following properties are equivalent:*

1.  $L \in \text{UPol}(\text{BPol}(\mathcal{G}^+))$ .
2.  $L \in \text{Pol}(\text{BPol}(\mathcal{G}^+)) \cap \text{coPol}(\text{BPol}(\mathcal{G}^+))$ .
3.  $L \in \Delta_2(<, +1, \mathbb{P}_{\mathcal{G}})$ .
4.  $L \in \text{FO}^2(<, +1, \mathbb{P}_{\mathcal{G}})$ .
5.  $L \in \text{TLX}(\mathcal{G})$ .
6. *The strict  $\mathcal{G}$ -kernel of the syntactic morphism of  $L$  belongs to **LDA**.*

**PROOF.** In this case as well, we already proved most of the implications. By Lemma 3.3, we have  $\Delta_2(<, +1, \mathbb{P}_{\mathcal{G}}) = \Delta_2(\mathbb{I}_{\mathcal{G}^+})$ . Hence, (2)  $\Leftrightarrow$  (3) follows from Theorem 4.9. Moreover, (1)  $\Leftrightarrow$  (2) is given by Theorem 4.21. Finally, since  $\text{FO}^2(<, +1, \mathbb{P}_{\mathcal{G}}) = \text{FO}^2(\mathbb{I}_{\mathcal{G}^+})$  by Lemma 3.3 and  $\text{TLX}(\mathcal{G}) = \text{TL}(\mathcal{G}^+)$  by Lemma 9.2, the equivalence (4)  $\Leftrightarrow$  (5) is immediate from Theorem 9.9. Finally, the implication (1)  $\Rightarrow$  (5) follows from Proposition 9.12. We now prove independently that (6)  $\Rightarrow$  (2) and (5)  $\Rightarrow$  (6) to complete the proof. We let  $\alpha : A^* \rightarrow M$  be the syntactic morphism of  $L$ . Let also  $N$  be the  $\mathcal{G}$ -kernel of  $\alpha$ , and  $S$  be its strict  $\mathcal{G}$ -kernel. Recall from Section 5 that we have  $S = N \cap \alpha(A^+)$  by definition.

We start with the implication (6)  $\Rightarrow$  (2). Assume that  $S$  belongs to **LDA**. We have to show that  $L \in \text{Pol}(\text{BPol}(\mathcal{G}^+)) \cap \text{coPol}(\text{BPol}(\mathcal{G}^+))$ . Since  $L$  is recognized by its syntactic morphism  $\alpha$ , it suffices to show that  $\alpha$  is a  $(\text{Pol}(\text{BPol}(\mathcal{G}^+)) \cap \text{coPol}(\text{BPol}(\mathcal{G}^+)))$ -morphism. By Theorem 6.7, this boils down to proving that for every  $s, t \in M$  and  $e \in E(M)$  such that  $(e, s) \in M^2$  is a  $\mathcal{G}^+$ -pair, we have  $(eset)^{\omega+1} = (eset)^\omega et(eset)^\omega$ . By Fact 5.6, we have  $e \in N$ . We consider two cases depending on whether  $e \in S$  or  $e \in N \setminus S$ . Assume first that  $e \in N \setminus S$ . Since  $S = N \cap \alpha(A^+)$ , this implies that  $e = 1_M$  and that  $\alpha^{-1}(1_M) = \{\varepsilon\}$ . Since  $(1_M, s)$  is a  $\mathcal{G}^+$ -pair and  $\{\varepsilon\} \in \mathcal{G}^+$ , it follows that  $s = 1_M$  as well. Hence,  $(eset)^{\omega+1} = t^{\omega+1} = (eset)^\omega et(eset)^\omega$  follows directly. We now consider the case where  $e \in S$ . By Lemma 5.13, we have  $es$  and  $et(eset)^{2\omega-1} \in N$ . Since  $e \in S$ , this implies that  $es \in S$  and  $et(eset)^{2\omega-1} \in S$ . Since  $S$  belongs to **LDA**, we get from Equation (16) that,

$$(eeseet(eset)^{2\omega-1}e)^\omega = (eeseet(eset)^{2\omega-1}e)^\omega eet(eset)^{2\omega-1}e(eeseet(eset)^{2\omega-1}e)^\omega.$$

Since  $e$  is idempotent, this simplifies into  $(eset)^\omega e = (eset)^\omega et(eset)^{3\omega-1}e$ . It now suffices to multiply by  $set$  on the right to get the desired equality  $(eset)^{\omega+1} = (eset)^\omega et(eset)^\omega$ .

We finally turn to the implication (5)  $\Rightarrow$  (6). Assume that  $L \in \text{TLX}(\mathcal{G}) = \text{TL}(\mathcal{G}^+)$ . We show that  $S$  belongs to **LDA**, i.e., that  $S$  satisfies (16). Let  $s, t \in S$  and  $e \in E(S)$ . We prove that  $(esete)^\omega = (esete)^\omega ete(esete)^\omega$ . By hypothesis,  $L$  is defined by a formula  $\varphi_L$  of  $\text{TL}[\mathcal{G}^+]$ . Let  $k \in \mathbb{N}$  be the rank of  $\varphi$ . Fact 9.3 yields a  $\mathcal{G}^+$ -morphism  $\eta : A^* \rightarrow Q$  such that  $\varphi_L$  is a  $\text{TL}[\eta]$  formula. It follows from Lemma 2.15 that  $G = \eta(A^+)$  is a group in  $N$  and that the morphism  $\beta : A^* \rightarrow G$  defined by  $\beta(w) = \eta(w)$  for every  $w \in A^+$  is a  $\mathcal{G}$ -morphism. Using the hypothesis that  $s, t, e \in S$ , we prove that there exist  $u, v, z \in A^*$  such that  $\alpha(u) = s$ ,  $\alpha(v) = t$ ,  $\alpha(z) = e$  and  $\eta(u) = \eta(v) = \eta(z) = 1_G$ . By symmetry, we only prove the existence of  $u \in A^*$ . There are two cases, depending on whether  $s = 1_M$  or not. If  $s = 1_M$ , then we have  $1_M \in S \subseteq \alpha(A^+)$ . Thus, there exists  $u' \in A^+$  such that  $\alpha(u') = 1_M$ . We let  $u = (u')^p$  for  $p = \omega(G)$ . Clearly,  $\alpha(u) = 1_M = s$  and  $\eta(u) = (\eta(u'))^p = 1_G$  since  $G$  is a group. Assume now that  $s \neq 1_M$ . We have  $s \in S \subseteq N$ . Since  $\beta : A^* \rightarrow G$  is a  $\mathcal{G}$ -morphism, it follows from Lemma 5.4 that there exists  $u \in A^*$  such that  $\alpha(u) = s$  and  $\beta(u) = 1_G$ . Since  $s \neq 1_M$ , we have  $u \in A^+$ , which also implies that  $\eta(u) = \beta(u) = 1_G$ , as desired.

We now use  $u, v, z$  (satisfying  $\alpha(u) = s$ ,  $\alpha(v) = t$ ,  $\alpha(z) = e$  and  $\eta(u) = \eta(v) = \eta(z) = 1_G$ ) to conclude the proof. Since  $\varphi_L$  is an  $\eta$ -formula of rank  $k$  and  $1_G \in E(N)$ , it follows from Lemma 9.6 and Proposition 9.7 that for all  $x, y \in A^*$ , we have,

$$x(z^k uz^{2k} vz^k)^k (z^k uz^{2k} vz^k)^k y \in L \Leftrightarrow x(z^k uz^{2k} vz^k)^k z^k vz^k (z^k uz^{2k} vz^k)^k y \in L.$$

Hence,  $(z^k uz^{2k} vz^k)^k (z^k uz^{2k} vz^k)^k$  and  $(z^k uz^{2k} vz^k)^k z^k vz^k (z^k uz^{2k} vz^k)^k$  are equivalent for the syntactic congruence of  $L$ . Since  $\alpha$  is the syntactic morphism of  $L$ , these words have the same image under  $\alpha$ . This yields  $(esete)^k (esete)^k = (esete)^k ete(esete)^k$  by definition of  $u, v$  and  $z$ . It now suffices to multiply by enough copies of  $esete$  on the left and on the right to obtain  $(esete)^\omega = (esete)^\omega ete(esete)^\omega$ , as desired. ■

Given as input a regular language  $L$ , one can compute its syntactic morphism. Moreover, by Lemma 5.3, the strict  $\mathcal{G}$ -kernel of this morphism can be computed as soon as  $\mathcal{G}$ -separation is decidable. It is then simple to decide whether it belongs to **LDA**: this boils down to checking if (16) holds by testing all possible combinations for  $s$ ,  $t$  and  $e$ . By Theorem 10.4, this decides whether  $L \in \text{UPol}(\text{BPol}(\mathcal{G}^+))$ . We state this in the following corollary.

**COROLLARY 10.5.** Let  $\mathcal{G}$  be a prevariety of group languages with decidable separation. Then, membership is decidable for  $\text{FO}^2(<, +1, \mathbb{P}_{\mathcal{G}}) = \text{TLX}(\mathcal{G}) = \Delta_2(<, +1, \mathbb{P}_{\mathcal{G}}) = \text{UPol}(\text{BPol}(\mathcal{G}^+))$ .

Recall again that separation is decidable for the prevarieties of group languages ST, MOD, AMT and GR, so that Corollary 10.5 applies in these cases: membership is decidable for the logical classes  $\text{FO}^2(<, +1)$ ,  $\text{FO}^2(<, +1, \text{MOD})$ ,  $\text{FO}^2(<, +1, \text{AMOD})$  and  $\text{FO}^2(<, +1, \mathbb{P}_{\text{GR}})$ .

Here again, Theorem 10.4 generalizes known results in the particular case  $\mathcal{G} = \text{ST}$ . Since  $\text{ST} = \{\emptyset, A^*\}$ , it is straightforward to verify that the strict ST-kernel of a morphism  $\alpha : A^* \rightarrow M$  is the semigroup  $\alpha(A^+)$ . When  $\alpha$  is the syntactic morphism of a language  $L$ , this object is called the *syntactic semigroup* of  $L$ . Thus, Theorem 10.4 yields that  $\Delta_2(<, +1) = \text{FO}^2(<, +1) = \text{F} + \text{X} + \text{P} + \text{Y}$  and that a regular language belongs to this class if and only if its syntactic semigroup belongs to **LDA**. As seen in the previous section, the equality  $\text{FO}^2(<, +1) = \text{F} + \text{X} + \text{P} + \text{Y}$  is due to Etessami, Vardi and Wilke [9]. The remaining correspondences are due to Thérien and Wilke [39]. In particular, the connection with **LDA** is established by relying on results of Almeida [1], based on a complex algebraic framework by Tilson [42], which involves categories and wreath products of finite semigroups. We bypass this intricate machinery here.

**REMARK 10.6.** In this case as well, there exists no generic result concerning separation and covering for classes of the form  $\text{UPol}(\text{BPol}(\mathcal{G}^+))$ . However, both problems are known to be decidable in *two* particular cases:  $\mathcal{G} = \text{ST}$  and  $\mathcal{G} = \text{MOD}$ . In view of Theorem 10.4, this corresponds to the logical classes  $\text{FO}^2(<, +1)$  and  $\text{FO}^2(<, +1, \text{MOD})$ . The decidability of covering for these two classes can be obtained by applying generic transfer results of [25] and [23] to the simpler class  $\text{FO}^2(<)$  (as we mentioned above, decidability of covering for this class was first proved in [32]). The techniques involved in these results are orthogonal and independent from the ones of the present paper.

## 11. Conclusion

We presented a generic algebraic characterization of unambiguous polynomial closure and used it to prove that when  $C$  is a prevariety, membership for  $\text{UPol}(C)$  reduces to membership for  $C$ . An interesting byproduct of the proof is that  $\text{UPol}(C) = \text{APol}(C) = \text{Pol}(C) \cap \text{coPol}(C)$  in that case. Moreover, we showed that when  $C$  is a finite prevariety, covering and separation are decidable for  $\text{UPol}(C)$ . This completes similar results of [31] for  $\text{Pol}(C)$  and  $\text{Bool}(\text{Pol}(C))$

and of [22] for  $Pol(Bool(Pol(C)))$ . Finally, we presented logical characterizations of the classes built with unambiguous polynomial closure. In particular, we proved that if the input class  $\mathcal{G}$  is a prevariety of *group languages*, then  $UPol(BPol(\mathcal{G})) = \Delta_2(<, \mathbb{P}_{\mathcal{G}}) = FO^2(<, \mathbb{P}_{\mathcal{G}}) = TL(\mathcal{G})$  and  $UPol(BPol(\mathcal{G}^+)) = \Delta_2(<, +1, \mathbb{P}_{\mathcal{G}}) = FO^2(<, +1, \mathbb{P}_{\mathcal{G}}) = TLX(\mathcal{G})$ . This generalizes earlier results corresponding to particular examples of prevarieties of group languages  $\mathcal{G}$ .

A natural follow-up question is whether our result for separation and covering can be pushed further to encompass more inputs than the finite classes. In particular, in view of the above logical characterizations, it would be interesting to look at both problems for classes of the form  $UPol(BPol(\mathcal{G}))$  and  $UPol(BPol(\mathcal{G}^+))$  when  $\mathcal{G}$  is a prevariety of group languages. While this is difficult, there are known results of this kind; in particular, it is known [30] that separation and covering are decidable for  $Pol(BPol(\mathcal{G}))$  as soon as separation is decidable for  $\mathcal{G}$ .

---

## References

- [1] Jorge Almeida. A syntactical proof of locality of DA. *International Journal of Algebra and Computation*, 6(2):165–177, 1996 [DOI](#) (72).
- [2] Jorge Almeida, Jana Bartonová, Ondrej Klíma, and Michal Kunc. On decidability of intermediate levels of concatenation hierarchies. *Proceedings of the 19th International Conference on Developments in Language Theory, DLT'15*, volume 9168 of *Lect. Notes Comp. Sci.* Pages 58–70. Springer, 2015 [DOI](#) (3, 4, 34).
- [3] Mustapha Arfi. Opérations polynomiales et hiérarchies de concaténation. *Theoretical Computer Science*, 91(1):71–84, 1991 [DOI](#) (2, 17, 18).
- [4] Mustapha Arfi. Polynomial operations on rational languages. *Proceedings of the 4th Annual Symposium on Theoretical Aspects of Computer Science, STACS'87*, volume 247 of *Lect. Notes Comp. Sci.* Pages 198–206. Springer, 1987 [DOI](#) (2, 17, 18).
- [5] Christopher J. Ash. Inevitable graphs: a proof of the type II conjecture and some related decision procedures. *International Journal of Algebra and Computation*, 1(1):127–146, 1991 [DOI](#) (9).
- [6] Mário Branco and Jean-Éric Pin. Equations defining the polynomial closure of a lattice of regular languages. *36th International Colloquium on Automata, Languages, and Programming, ICALP'09*, volume 5556 of *Lect. Notes Comp. Sci.* Pages 115–126. Springer, 2009 [DOI](#) (3).
- [7] Luc Dartois and Charles Paperman. Two-variable first order logic with modular predicates over words. *Proceedings of the 30th International Symposium on Theoretical Aspects of Computer Science, STACS'13*, volume 20 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 329–340. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2013 [DOI](#) (4, 70).
- [8] Volker Diekert, Paul Gastin, and Manfred Kufleitner. A survey on small fragments of first-order logic over finite words. *International Journal of Foundations of Computer Science*, 19(3):513–548, 2008 [DOI](#) (3).
- [9] Kousha Etessami, Moshe Y. Vardi, and Thomas Wilke. First-order logic with two variables and unary temporal logic. *Information and Computation*, 179(2):279–295, 2002 [DOI](#) (3, 54, 61, 69, 72).
- [10] James Alexander Green. On the structure of semigroups. *Annals of Mathematics*, 54(1):163–172, 1951 [DOI](#) (6).
- [11] Hans W. Kamp. Tense Logic and the Theory of Linear Order. PhD thesis, Computer Science Department, University of California at Los Angeles, USA, 1968 (1).
- [12] Andreas Krebs, Kamal Lodaya, Paritosh K. Pandya, and Howard Straubing. Two-variable logics with some betweenness relations: expressiveness, satisfiability and membership. *Logical Methods in Computer Science*, 16(3), 2020 [DOI](#) (65).
- [13] Manfred Kufleitner and Tobias Walter. One quantifier alternation in first-order logic with modular predicates. *RAIRO Theoretical Informatics and Applications*, 49(1):1–22, 2015 [DOI](#) (4, 70).
- [14] Robert McNaughton and Seymour A. Papert. Counter-Free Automata. MIT Press, 1971 (1).
- [15] Jean-Éric Pin. Propriétés syntactiques du produit non ambigu. *Proceedings of the 7th International Colloquium on Automata, Languages and Programming, ICALP'80*, volume 85 of *Lect. Notes Comp. Sci.* Pages 483–499. Springer, 1980 [DOI](#) (3, 38).

- [16] Jean-Eric Pin. An explicit formula for the intersection of two polynomials of regular languages. *Proceedings of the 17th International Conference on Developments in Language Theory, DLT'13*, volume 7907 of *Lect. Notes Comp. Sci.* Pages 31–45. Springer, 2013 [DOI](#) (18, 22).
- [17] Jean-Eric Pin. Mathematical foundations of automata theory. In preparation, 2023 [URL](#) (7).
- [18] Jean-Eric Pin. The dot-depth hierarchy, 45 years later. *The Role of Theory in Computer Science. Essays Dedicated to Janusz Brzozowski*, chapter 8, pages 177–202. World Scientific, 2017 [DOI](#) (2).
- [19] Jean-Eric Pin, Howard Straubing, and Denis Thérien. Locally trivial categories and unambiguous concatenation. *Journal of Pure and Applied Algebra*, 52(3):297–311, 1988 [DOI](#) (3, 4, 23).
- [20] Jean-Eric Pin and Pascal Weil. Polynomial closure and unambiguous product. *Theory of Computing Systems*, 30(4):383–422, 1997 [DOI](#) (2–4, 22, 69).
- [21] Jean-Éric Pin and Howard Straubing. Monoids of upper triangular boolean matrices, *Semigroups. Structure and Universal Algebraic Problems*. Volume 39, *Colloquia Mathematica Societatis Janos Bolyai*, pages 259–272. North-Holland, 1985 [URL](#) (18).
- [22] Thomas Place. Separating regular languages with two quantifier alternations. *Logical Methods in Computer Science*, 14(4), 2018 [DOI](#) (2, 3, 34, 73).
- [23] Thomas Place, Varun Ramanathan, and Pascal Weil. Covering and separation for logical fragments with modular predicates. *Logical Methods in Computer Science*, 15(2), 2019 [DOI](#) (72).
- [24] Thomas Place, Lorijn van Rooijen, and Marc Zeitoun. Separating regular languages by piecewise testable and unambiguous languages. *Proceedings of the 38th International Symposium on Mathematical Foundations of Computer Science, MFCS'13*, volume 8087 of *Lect. Notes Comp. Sci.* Pages 729–740. Springer, 2013 [DOI](#) (3).
- [25] Thomas Place and Marc Zeitoun. Adding successor: A transfer theorem for separation and covering. *ACM Transactions on Computational Logic*, 21(2):9:1–9:45, 2020 [DOI](#) (72).
- [26] Thomas Place and Marc Zeitoun. Generic results for concatenation hierarchies. *Theory of Computing Systems (ToCS)*, 63:849–901, 2019 [DOI](#) (2, 4, 5, 18, 19, 33).
- [27] Thomas Place and Marc Zeitoun. Going higher in first-order quantifier alternation hierarchies on words. *Journal of the ACM*, 66(2):12:1–12:65, 2019 [DOI](#) (2).
- [28] Thomas Place and Marc Zeitoun. Group separation strikes back. *38th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS'23*, pages 1–13. IEEE Computer Society, 2023 [DOI](#) (9).
- [29] Thomas Place and Marc Zeitoun. Separating without any ambiguity. *45th International Colloquium on Automata, Languages, and Programming, ICALP'18*, Leibniz International Proceedings in Informatics (LIPIcs), 137:1–137:14. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2018 [DOI](#) (5).
- [30] Thomas Place and Marc Zeitoun. Separation and covering for group based concatenation hierarchies. *Proceedings of the 34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS'19*, pages 1–13. IEEE Computer Society, 2019 [DOI](#) (2, 73).
- [31] Thomas Place and Marc Zeitoun. Separation for dot-depth two. *Logical Methods in Computer Science*, Volume 17, Issue 3, 2021 [DOI](#) (2, 72).
- [32] Thomas Place and Marc Zeitoun. The covering problem. *Logical Methods in Computer Science*, 14(3), 2018 [DOI](#) (8, 44–47, 49, 72).
- [33] John L. Rhodes. A homomorphism theorem for finite semigroups. *Mathematical systems theory*, 1:289–304, 1967 [DOI](#) (3).
- [34] Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8(2):190–194, 1965 [DOI](#) (1).
- [35] Marcel Paul Schützenberger. Sur le produit de concaténation non ambigu. *Semigroup Forum*, 13:47–75, 1976 [DOI](#) (3, 22, 68, 69).
- [36] Imre Simon. Piecewise testable events. *Proceedings of the 2nd GI Conference on Automata Theory and Formal Languages*, pages 214–222. Springer, 1975 [DOI](#) (2, 18).
- [37] Howard Straubing. On logical descriptions of regular languages. *LATIN 2002: Theoretical Informatics, 5th Latin American Symposium*, volume 2286 of *Lect. Notes Comp. Sci.* Pages 528–538. Springer, 2002 [DOI](#) (26).
- [38] Pascal Tesson and Denis Thérien. Diamonds are forever: the variety DA. *Semigroups, Algorithms, Automata and Languages*, pages 475–500. World Scientific, 2002 [DOI](#) (3, 68).
- [39] Denis Thérien and Thomas Wilke. Over words, two variables are as powerful as one quantifier alternation. *Proceedings of the 30th Annual ACM Symposium on Theory of Computing, STOC'98*, pages 234–240, New York, NY, USA. ACM, 1998 [DOI](#) (3, 4, 68, 69, 72).
- [40] Gabriel Thierrin. Permutation automata. *Theory of Computing Systems*, 2(1):83–90, 1968 [DOI](#) (4).
- [41] Wolfgang Thomas. Classifying regular events in symbolic logic. *Journal of Computer and System Sciences*, 25(3):360–376, 1982 [DOI](#) (2, 19).
- [42] Bret Tilson. Categories as algebra: essential ingredient in the theory of monoids. *Journal of Pure and Applied Algebra*, 48(1):83–198, 1987 [DOI](#) (72).