A Generic Characterization of Generalized Unary Temporal Logic and Two-Variable First-Order Logic

Thomas Place
Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France

Marc Zeitoun
Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400, Talence, France

Abstract

We study an operator on classes of languages. For each class $\mathcal{C}$, it produces a new class $\mathrm{FO}^2(\mathbb{I}_{\mathcal{C}})$ associated with a variant of two-variable first-order logic equipped with a signature $\mathbb{I}_{\mathcal{C}}$ built from $\mathcal{C}$. For $\mathcal{C} = \{\emptyset, A^*\}$, we obtain the usual $\mathrm{FO}^2(<)$ logic, equipped with linear order. For $\mathcal{C} = \{\emptyset, \{\varepsilon\}, A^+, A^*\}$, we get the variant $\mathrm{FO}^2(<, +1)$, which also includes the successor predicate. If $\mathcal{C}$ consists of all Boolean combinations of languages A^*aA^*, where a is a letter, we get the variant $\mathrm{FO}^2(<, \mathit{Bet})$, which includes "between" relations. We prove a *generic* algebraic characterization of the classes $\mathrm{FO}^2(\mathbb{I}_{\mathcal{C}})$. It elegantly generalizes those known for all the cases mentioned above. Moreover, it implies that if $\mathcal{C}$ has decidable separation (plus some standard properties), then $\mathrm{FO}^2(\mathbb{I}_{\mathcal{C}})$ has a decidable membership problem.

We actually work with an equivalent definition of $\mathrm{FO}^2(\mathbb{I}_{\mathcal{C}})$ in terms of *unary temporal logic*. For each class $\mathcal{C}$, we consider a variant $\mathrm{TL}(\mathcal{C})$ of unary temporal logic whose future/past modalities depend on $\mathcal{C}$ and such that $\mathrm{TL}(\mathcal{C}) = \mathrm{FO}^2(\mathbb{I}_{\mathcal{C}})$. Finally, we also characterize $\mathrm{FL}(\mathcal{C})$ and $\mathrm{PL}(\mathcal{C})$, the pure-future and pure-past restrictions of $\mathrm{TL}(\mathcal{C})$. Like for $\mathrm{TL}(\mathcal{C})$, these characterizations imply that if $\mathcal{C}$ is a class with decidable separation, then $\mathrm{FL}(\mathcal{C})$ and $\mathrm{PL}(\mathcal{C})$ have decidable membership.

2012 ACM Subject Classification Theory of computation → Formal languages and automata theory

Keywords and phrases Classes of regular languages, Generalized unary temporal logic, Generalized two-variable first-order logic, Generic decidable characterizations, Membership, Separation

Digital Object Identifier 10.4230/LIPIcs.CSL.2024.45

Related Version *Extended Version*: https://arxiv.org/abs/2307.09349 [29]

1 Introduction

Context. Regular languages of finite words form a robust class: they admit a wide variety of equivalent definitions, whether by regular expressions, finite automata, finite monoids or monadic second-order logic. It is therefore natural to study the *fragments* of regular languages obtained by restricting the syntax of one of the above-mentioned formalisms. For each particular fragment, we seek to prove that it has a decidable *membership problem*: given a regular language as input, decide whether it belongs to the fragment. Intuitively, doing so requires a thorough knowledge of the fragment and the languages it can describe.

This approach was initiated by Schützenberger [30] for the class SF of star-free languages. These are the languages defined by a *star-free expression*: a regular expression without Kleene star but with complement instead. Equivalently, these are the languages that can be defined in first-order logic with the linear order [16] (FO(<)) or in linear temporal logic [11] (LTL). Schützenberger established an algebraic characterization of SF: a regular language is star-free if and only if its *syntactic monoid is aperiodic*. This yields a membership procedure for SF because the syntactic monoid can be computed and aperiodicity is a decidable property.

Operators. This seminal result prompted researchers to look at other natural classes, spawning a fruitful line of research (see *e.g.*, [4, 33, 12, 18, 36, 24]). Although there are numerous classes, they can be grouped into families based on "variants" of the same syntax. Let us

32nd EACSL Annual Conference on Computer Science Logic (CSL 2024).
Editors: Aniello Murano and Alexandra Silva; Article No. 45; pp. 45:1–45:23
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

use logic to clarify this point. Each logical fragment can use several *signatures* (*i.e.*, sets of predicates allowed in formulas), each giving rise to a class. For instance, first-order logic is commonly equipped with predicates such as the linear order "$<$" [16, 30], the successor "$+1$" [3] or the modular predicates "MOD" [2]. While it is worth looking at multiple variants of prominent classes, doing so individually for each of them has an obvious disadvantage: the proof has to be systematically modified to accommodate each change. This can be tedious, difficult, and not necessarily enlightening. To overcome this drawback, a natural approach is to capture a whole family of variants with an *operator*. An operator "Op" takes a class $\mathcal{C}$ as input, and outputs a larger one $\mathrm{Op}(\mathcal{C})$. Thus, we can study all classes $\mathrm{Op}(\mathcal{C})$ *simultaneously*: the question becomes: "what hypotheses about $\mathcal{C}$ guarantee the decidability of $\mathrm{Op}(\mathcal{C})$-membership?". For example, one can generalize the three definitions of star-free languages through operators:

1. The *star-free closure* $\mathcal{C} \mapsto \mathrm{SF}(\mathcal{C})$ has been introduced in [31, 34]. Languages in $\mathrm{SF}(\mathcal{C})$ are defined by "extended" star-free expressions, which can freely use languages from $\mathcal{C}$.
2. A construction associating a signature $\mathbb{I}_{\mathcal{C}}$ to a class $\mathcal{C}$ has been given in [23]. For each $L \in \mathcal{C}$, the set $\mathbb{I}_{\mathcal{C}}$ contains a binary predicate $I_L(x, y)$: for a word w and two positions i, j in w, $I_L(i, j)$ holds if and only if $i < j$ and the infix of w between i and j belongs to L. We get an operator $\mathcal{C} \mapsto \mathrm{FO}(\mathbb{I}_{\mathcal{C}})$ based on first-order logic. It captures many choices of signature.
3. Similarly, an operator $\mathcal{C} \mapsto \mathrm{LTL}(\mathcal{C})$ that generalizes LTL has been defined in [28].

It is shown in [23, 28] that $\mathrm{SF}(\mathcal{C}) = \mathrm{FO}(\mathbb{I}_{\mathcal{C}}) = \mathrm{LTL}(\mathcal{C})$ for any class $\mathcal{C}$ (with mild hypotheses). Moreover, a *generic* algebraic characterization is proved in [25, 28]. Given a regular language L, it relies on a construction that identifies monoids inside its syntactic monoid, called the *$\mathcal{C}$-orbits*: $L \in \mathrm{SF}(\mathcal{C})$ if and only if its $\mathcal{C}$-orbits are all *aperiodic*. This elegantly generalizes Schützenberger's theorem and gives a *transfer theorem* for membership. Indeed, the $\mathcal{C}$-orbits are connected with a decision problem that strengthens membership: *$\mathcal{C}$-separation*. Given *two* input regular languages L_1 and L_2, $\mathcal{C}$-separation asks whether there is $K \in \mathcal{C}$ such that $L_1 \subseteq K$ and $L_2 \cap K = \emptyset$. The crucial point is that $\mathcal{C}$-orbits are computable if $\mathcal{C}$-separation is decidable. Thus, $\mathrm{SF}(\mathcal{C})$-membership is also decidable in this case. Similar results are known for other operators such as *polynomial closure* [23] or its *unambiguous* restriction [22, 27].

Unary temporal logic and two-variable first-order logic. The operator we investigate generalizes another important class admitting multiple definitions [35, 6] (see [8, 7] for extensions). We are interested in two of them. It consists of languages that can be defined in *two-variable first-order logic* with the linear order ($\mathrm{FO}^2(<)$) or equivalently in *unary temporal logic* (TL) with the modalities F (sometimes in the future) and P (sometimes in the past). Etessami, Vardi and Wilke [9] have shown that $\mathrm{FO}^2(<) = \mathrm{TL}$. Its algebraic characterization by Thérien and Wilke [36] is one of the famous results of this type: a regular language belongs to $\mathrm{FO}^2(<) = \mathrm{TL}$ if and only if its syntactic monoid belongs to the variety of monoids DA.

Both definitions extend to natural operators. First, the generic signatures $\mathbb{I}_{\mathcal{C}}$ yield an operator $\mathcal{C} \mapsto \mathrm{FO}^2(\mathbb{I}_{\mathcal{C}})$ based on FO^2. Second, an operator $\mathcal{C} \mapsto \mathrm{TL}(\mathcal{C})$ has been defined in [27]. It enriches TL with new modalities F_L and P_L, both depending on the languages $L \in \mathcal{C}$. For example, the formula $\mathrm{F}_L\, \varphi$ holds at a position i in a word w if there is a position $j > i$ in w such that φ holds at j and the infix between i and j belongs to L. We know that $\mathrm{FO}^2(\mathbb{I}_{\mathcal{C}}) = \mathrm{TL}(\mathcal{C})$ when $\mathcal{C}$ is closed under Boolean operations [27]. Here, we work with the $\mathrm{TL}(\cdot)$ operator, which encompasses all classic classes based on two-variable first-order logic or unary temporal logic. This includes the original variants $\mathrm{FO}^2(<) = \mathrm{TL}$ and $\mathrm{FO}^2(<, +1) = \mathrm{TLX}$, both of which were studied by Thérien and Wilke [36] (here, "$+1$" is the successor predicate and TLX is defined by enriching TL with "next" and "yesterday" modalities). Another example is the variant

$\mathrm{FO}^2(<, MOD)$ endowed with *modular predicates*, investigated by Dartois and Paperman [5]. Finally, we capture the variant $\mathrm{FO}^2(<, Bet) = \mathrm{BInvTL}$ equipped with "*between*" relations, defined and characterized by Krebs, Lodaya, Pandya and Straubing [13, 14, 15].

Contributions. We prove a *generic* algebraic characterization of the classes $\mathrm{FO}^2(\mathbb{I}_{\mathcal{C}}) = \mathrm{TL}(\mathcal{C})$. We reuse the $\mathcal{C}$-orbits introduced for star-free closure: for any class $\mathcal{C}$ (having mild properties) we show that a regular language belongs to $\mathrm{TL}(\mathcal{C})$ if and only if all $\mathcal{C}$-orbits of its syntactic monoid belong to DA. In particular, this yields a *transfer theorem* for membership: if $\mathcal{C}$ has decidable *separation*, then $\mathrm{FO}^2(\mathbb{I}_{\mathcal{C}}) = \mathrm{TL}(\mathcal{C})$ has decidable membership. Moreover, this characterization generalizes the characterizations known for all the above instances.

A key feature of our proof is that we use a third auxiliary operator. It combines two other operators: Boolean polynomial closure (BPol) and unambiguous polynomial closure (UPol). We have $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C})) \subseteq \mathrm{TL}(\mathcal{C})$ if $\mathcal{C}$ has mild properties [27]. In fact, for many natural classes, $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C})) = \mathrm{TL}(\mathcal{C})$. For example, $\mathrm{UPol}(\mathrm{BPol}(\{\emptyset, A^*\}))$ is the class UL of *unambiguous languages* defined by Schützenberger [32]. It is known [36] that $\mathrm{UL} = \mathrm{TL} = \mathrm{FO}^2(<)$. More generally, $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C})) = \mathrm{TL}(\mathcal{C})$ for every class $\mathcal{C}$ consisting of *group languages* [27]. Yet, this is a strong hypothesis and the inclusion $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C})) \subseteq \mathrm{TL}(\mathcal{C})$ is *strict* in general. For example, the results of [15] imply that $\mathrm{UPol}(\mathrm{BPol}(\mathrm{AT})) \neq \mathrm{TL}(\mathrm{AT})$, where AT consists of all Boolean combinations of languages A^*aA^* (with $a \in A$). Nevertheless, the classes $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C}))$ serve as a central ingredient in the most difficult direction of our proof: "*If a language satisfies our characterization on $\mathcal{C}$-orbits, prove that it belongs to* $\mathrm{TL}(\mathcal{C})$". More precisely, we exploit the known characterization of $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C}))$ to prove that auxiliary languages belong to this class, and we then conclude using the inclusion $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C})) \subseteq \mathrm{TL}(\mathcal{C})$.

Finally, we look at two additional operators: $\mathcal{C} \mapsto \mathrm{FL}(\mathcal{C})$ and $\mathcal{C} \mapsto \mathrm{PL}(\mathcal{C})$. They are also defined in terms of unary temporal logic, as the pure-future and the pure-past restrictions of $\mathcal{C} \mapsto \mathrm{TL}(\mathcal{C})$. We present generic algebraic characterizations for these two operators as well. Again, they are based on $\mathcal{C}$-orbits. For every class $\mathcal{C}$ (with mild hypotheses), we show that a regular language belongs to $\mathrm{FL}(\mathcal{C})$ (resp. $\mathrm{PL}(\mathcal{C})$) if and only if all the $\mathcal{C}$-orbits inside its syntactic monoid are $\mathcal{L}$-trivial (resp. $\mathcal{R}$-trivial) monoids. As before, these results yield *transfer theorems*: if $\mathcal{C}$ has decidable *separation*, then $\mathrm{FL}(\mathcal{C})$ and $\mathrm{PL}(\mathcal{C})$ have decidable membership.

Organization of the paper. We recall the notation and background in Section 2. In Section 3, we present the $\mathcal{C}$-orbits and their properties. In Section 4, we define the operator $\mathcal{C} \mapsto \mathrm{TL}(\mathcal{C})$. Section 5 is devoted to the generic characterization of $\mathrm{TL}(\mathcal{C})$ and to its proof. In Section 6, finally, we state the characterizations of the pure-future and pure-past restrictions of $\mathrm{TL}(\mathcal{C})$.

2 Preliminaries

We fix a finite *alphabet* A for the paper. As usual, A^* denotes the set of all finite words over A, including the empty word ε. A *language* is a subset of A^*. We let $A^+ = A^* \setminus \{\varepsilon\}$. For $u, v \in A^*$, we write uv for the word obtained by concatenating u and v. We lift the concatenation to languages as follows: if $K, L \subseteq A^*$, we let $KL = \{uv \mid u \in K, v \in L\}$. If $w \in A^*$, we write $|w| \in \mathbb{N}$ for its length. A word $w = a_1 \cdots a_{|w|} \in A^*$ is viewed as an *ordered set* $\mathsf{Pos}(w) = \{0, 1, \dots, |w|, |w|+1\}$ *of* $|w|+2$ *positions*. In addition, we let $\mathsf{Pos}_c(w) = \{1, \dots, |w|\} \subsetneq \mathsf{Pos}(w)$. Position $i \in \mathsf{Pos}_c(w)$ carries label $a_i \in A$, which we write $w[i] = a_i$. On the other hand, positions 0 and $|w|+1$ carry *no label*. We write $w[0] = min$ and $w[|w|+1] = max$. For $v, w \in A^*$, we say that v is an *infix* (resp. *prefix*, *suffix*) of w when there exist $x, y \in A^*$ such that $w = xvy$ (resp. $w = vy$, $w = xv$). Given a word $w = a_1 \cdots a_{|w|} \in A^*$ and $i, j \in \mathsf{Pos}(w)$ 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 *strictly* between i and j).

Classes. A *class* of languages $\mathcal{C}$ is simply a set of languages. Such a class $\mathcal{C}$ is a *lattice* when $\emptyset \in \mathcal{C}$, $A^* \in \mathcal{C}$ and $\mathcal{C}$ is closed under both union and intersection: for all $K, L \in \mathcal{C}$, we have $K \cup L \in \mathcal{C}$ and $K \cap L \in \mathcal{C}$. Moreover, a class of languages $\mathcal{C}$ is a *Boolean algebra* if it is a lattice closed under complement: for all $L \in \mathcal{C}$, we have $A^* \setminus L \in \mathcal{C}$. Finally, the class $\mathcal{C}$ is *closed under quotients* if for all $L \in \mathcal{C}$ and $u \in A^*$, we have $u^{-1}L \stackrel{\text{def}}{=} \{w \in A^* \mid uw \in L\} \in \mathcal{C}$ and $Lu^{-1} \stackrel{\text{def}}{=} \{w \in A^* \mid wu \in L\} \in \mathcal{C}$. A *prevariety* is a Boolean algebra closed under quotients and containing only *regular languages*. Regular languages are those which can be equivalently defined by finite automata, finite monoids or monadic second-order logic. We work with the definition by monoids, which we now recall.

Monoids. A *monoid* is a set M endowed with an associative multiplication $(s, t) \mapsto st$ having an identity element 1_M (*i.e.*, such that $1_M s = s 1_M = s$ for every $s \in M$). An *idempotent* of a monoid M is an element $e \in M$ such that $ee = e$. We write $E(M) \subseteq M$ for the set of all idempotents in M. It is folklore that for every *finite* monoid M, there exists a natural number $\omega(M)$ (denoted by ω when M is understood) such that for every $s \in M$, the element s^ω is an idempotent. Finally, we shall use the following Green relations [10] defined on monoids. Given a monoid M and $s, t \in M$, we write:

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

Clearly, $\leqslant_{\mathcal{J}}$, $\leqslant_{\mathcal{L}}$ and $\leqslant_{\mathcal{R}}$ 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 $s \leqslant_{\mathcal{J}} t$ but $t \nleqslant_{\mathcal{J}} s$). Finally, we write $\mathcal{J}$, $\mathcal{L}$ and $\mathcal{R}$ for the corresponding equivalence relations (for example, $s \mathrel{\mathcal{J}} t$ when $s \leqslant_{\mathcal{J}} t$ and $t \leqslant_{\mathcal{J}} s$). There are many technical results about Green relations. We will just need the following easy and standard lemma, which applies to *finite* monoids (see *e.g.*, [17, 20]).

▶ **Lemma 1.** *Let M be a finite monoid and let $s, t \in M$. If $s \mathrel{\mathcal{J}} t$ and $s \leqslant_{\mathcal{R}} t$, then $s \mathrel{\mathcal{R}} t$.*

Regular languages and syntactic morphisms. Since A^* is a monoid whose multiplication is concatenation (the identity element is ε), we may consider monoid morphisms $\alpha : A^* \to M$ where M is an arbitrary monoid. That is, $\alpha : A^* \to M$ is a map satisfying $\alpha(\varepsilon) = 1_M$ and $\alpha(uv) = \alpha(u)\alpha(v)$ for all $u, v \in A^*$. We say that a language $L \subseteq A^*$ is *recognized* by α when there exists a set $F \subseteq M$ such that $L = \alpha^{-1}(F)$.

It is well known that regular languages are exactly those recognized by a morphism into a *finite* monoid. Moreover, every language L is recognized by a canonical morphism, which we briefly recall. One can associate to L an equivalence $\equiv_L$ over A^*: the *syntactic congruence of L*. Given $u, v \in A^*$, we let $u \equiv_L v$ if and only if $xuy \in L \Leftrightarrow xvy \in L$ for every $x, y \in A^*$. One can check that "$\equiv_L$" is indeed a congruence on A^*: it is an equivalence compatible with word concatenation. Thus, the set of equivalence classes $M_L = A^*/{\equiv_L}$ is a monoid. It is called the *syntactic monoid of L*. Finally, the map $\alpha_L : A^* \to M_L$ sending every word to its equivalence class is a morphism recognizing L, called the *syntactic morphism of L*. It is known that a language 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 $\alpha_L : A^* \to M_L$ from any representation of L (such as an automaton or a monoid morphism).

Decision problems. We consider two decision problems, both depending on an arbitrary class $\mathcal{C}$. They serve as mathematical tools for analyzing it, as obtaining an algorithm for one of these problems requires a solid understanding of that class $\mathcal{C}$. The *$\mathcal{C}$-membership problem* is the simplest: it takes as input a single regular language L and simply asks whether $L \in \mathcal{C}$. The second problem, *$\mathcal{C}$-separation*, is more general. Given three languages K, L_1, L_2, we say that K *separates* L_1 from L_2 if $L_1 \subseteq K$ and $L_2 \cap K = \emptyset$. Given a class $\mathcal{C}$, we say that L_1 is *$\mathcal{C}$-separable* from L_2 if some language of $\mathcal{C}$ separates L_1 from L_2. The *$\mathcal{C}$-separation problem* takes as input two regular languages L_1, L_2 and asks whether L_1 is $\mathcal{C}$-separable from L_2.

▶ Remark 2. The $\mathcal{C}$-separation problem generalizes $\mathcal{C}$-membership. Indeed, a regular language belongs to $\mathcal{C}$ if and only if it is $\mathcal{C}$-separable from its complement, which is regular.

3 Orbits

Instead of looking at single classes, we consider *operators*. These are correspondences $\mathcal{C} \mapsto \mathrm{Op}(\mathcal{C})$ that take as input a class $\mathcal{C}$ to build a new one $\mathrm{Op}(\mathcal{C})$. We investigate three operators in Sections 4 to 6. For now, we present general tools for handling such operators. Given a class $\mathcal{C}$ and a morphism $\alpha : A^* \to M$, we define special subsets of M: the $\mathcal{C}$-orbits for α. This notion was introduced in [28]. We shall use it to formulate *generic characterizations* of the operators $\mathcal{C} \mapsto \mathrm{Op}(\mathcal{C})$ that we consider: for each input prevariety $\mathcal{C}$, the languages in $\mathrm{Op}(\mathcal{C})$ are characterized by a property of the $\mathcal{C}$-orbits for their syntactic morphisms.

$\mathcal{C}$-pairs. Consider a class $\mathcal{C}$ and a morphism $\alpha : A^* \to M$. We say that a pair $(s,t) \in M^2$ is a *$\mathcal{C}$-pair for α* if and only if $\alpha^{-1}(s)$ is *not* $\mathcal{C}$-separable from $\alpha^{-1}(t)$. Note that if $\mathcal{C}$-separation is decidable, then one can compute all $\mathcal{C}$-pairs for an input morphism.

We turn to a useful technical result, which characterizes the $\mathcal{C}$-pairs using morphisms. Consider *two* morphisms $\alpha : A^* \to M$ and $\eta : A^* \to N$. For every pair $(s,t) \in M^2$, we say that (s,t) is an *η-pair for α* when there exist $u, v \in A^*$ such that $\eta(u) = \eta(v)$, $\alpha(u) = s$ and $\alpha(v) = t$. In addition, for each class $\mathcal{C}$, we define the *$\mathcal{C}$-morphisms* as the *surjective* morphisms $\eta : A^* \to N$ into a finite monoid N such that all languages recognized by η belong to $\mathcal{C}$. We have the following elementary lemma, proved in [27, Lemma 5.11].

▶ **Lemma 3.** *Let $\mathcal{C}$ be a prevariety and $\alpha : A^* \to M$ be a morphism. Then,*

1. *For every $\mathcal{C}$-morphism $\eta : A^* \to N$, all $\mathcal{C}$-pairs for α are also η-pairs for α.*
2. *There exists a $\mathcal{C}$-morphism $\eta : A^* \to N$ such that all η-pairs for α are also $\mathcal{C}$-pairs for α.*

$\mathcal{C}$-orbits. Consider a class $\mathcal{C}$ and a morphism $\alpha : A^* \to M$. For every *idempotent* $e \in E(M)$, the *$\mathcal{C}$-orbit of e for α* is the set $M_e \subseteq M$ consisting of all elements $ete \in M$ such that $(e,t) \in M^2$ is a $\mathcal{C}$-pair. If $\mathcal{C}$ is a prevariety and α is surjective, it is proved in [28, Lemma 5.5] that M_e is a monoid in M: it is closed under multiplication and $e \in M_e$ is its identity. On the other hand, M_e is not a "*sub*monoid" of M (this is because 1_M needs not belong to M_e).

▶ **Lemma 4.** *Let $\mathcal{C}$ be a prevariety and $\alpha : A^* \to M$ be a surjective morphism into a finite monoid. For all $e \in E(M)$, the $\mathcal{C}$-orbit of e for α is a monoid in M whose identity is e.*

As seen above, when $\mathcal{C}$ has decidable separation, one can compute the $\mathcal{C}$-pairs associated with an input morphism. Hence, one can also compute the $\mathcal{C}$-orbits in this case.

▶ **Lemma 5.** *Let $\mathcal{C}$ be a class with decidable separation. Given as input a morphism $\alpha : A^* \to M$ into a finite monoid and $e \in E(M)$, one can compute the $\mathcal{C}$-orbit of e for α.*

Finally, the following lemma connects $\mathcal{C}$-orbits with $\mathcal{C}$-morphisms.

▶ **Lemma 6.** *Let $\mathcal{C}$ be a prevariety and $\alpha : A^* \to M$ be a morphism. Moreover, let $\eta : A^* \to N$ be a $\mathcal{C}$-morphism. For every $e \in E(M)$, there exists $f \in E(N)$ such that the $\mathcal{C}$-orbit of e for α is contained in the set $\alpha(\eta^{-1}(f))$.*

Proof. Let $t_1, \dots, t_n \in M$ be all elements of the set $\{t \in M \mid (e,t) \text{ is a } \mathcal{C}\text{-pair}\}$. By definition, the $\mathcal{C}$-orbit of e for α is $M_e = \{et_1e, \dots, et_ne\}$. Since η is a $\mathcal{C}$-morphism, Lemma 3 implies that (e, t_i) is an η-pair for all $i \leq n$. This yields $x_i, y_i \in A^*$ such that $\eta(x_i) = \eta(y_i)$, $\alpha(x_i) = e$ and $\alpha(y_i) = t_i$. Let $p = \omega(N)$, $w = (x_1 \cdots x_n)^p$ and $f = \eta(w)$. Note that f is idempotent by choice of p. We show that $et_ie \in \alpha(\eta^{-1}(f))$ for $i \leq n$. We define $w_i = (x_1 \cdots x_n)^p x_1 \cdots x_{i-1} y_i x_{i+1} \cdots x_n (x_1 \cdots x_n)^{2p-1}$. By definition, we have $\alpha(w_i) = et_ie$. Now, since $\eta(x_i) = \eta(y_i)$, we get $\eta(w_i) = \eta(w) = f$. Hence, $et_ie \in \alpha(\eta^{-1}(f))$, as desired. ◀

4 Generalized unary temporal logic

In this section, we define generalized unary temporal logic. We introduce an *operator* $\mathcal{C} \mapsto \mathrm{TL}(\mathcal{C})$ that associates a new class of languages $\mathrm{TL}(\mathcal{C})$ with every input class $\mathcal{C}$. We first recall its definition (taken from [27]), and we then complete it with useful properties.

4.1 Definition

Syntax. We associate with any class $\mathcal{C}$ a set of temporal formulas denoted by $\mathrm{TL}[\mathcal{C}]$ as follows. A $\mathrm{TL}[\mathcal{C}]$ formula is built from atomic formulas using Boolean connectives and temporal operators. The atomic formulas are $\top$, $\bot$, min, max and "a" for every letter $a \in A$. All Boolean connectives are allowed: if ψ_1 and ψ_2 are $\mathrm{TL}[\mathcal{C}]$ formulas, then so are $(\psi_1 \vee \psi_2)$, $(\psi_1 \wedge \psi_2)$ and $(\neg\psi_1)$. We associate *two temporal modalities* with every language $L \in \mathcal{C}$, which we denote by F_L and P_L: if ψ is a $\mathrm{TL}[\mathcal{C}]$ formula, then so are $(\mathrm{F}_L\, \psi)$ and $(\mathrm{P}_L\, \psi)$. For the sake of improved readability, we omit parentheses when there is no ambiguity.

Semantics. Evaluating a $\mathrm{TL}[\mathcal{C}]$ formula φ requires a word $w \in A^*$ and a position $i \in \mathsf{Pos}(w)$. We define by induction what it means for (w, i) *to satisfy* φ, which one denotes by $w, i \models \varphi$.

- **Atomic formulas:** $w, i \models \top$ always holds, $w, i \models \bot$ never holds and for every symbol $\ell \in A \cup \{min, max\}$, $w, i \models \ell$ 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.
- **Finally:** for $L \in \mathcal{C}$, we let $w, i \models \mathrm{F}_L\, \psi$ when there exists $j \in \mathsf{Pos}(w)$ such that $i < j$, $w(i,j) \in L$ and $w, j \models \psi$.
- **Previously:** for $L \in \mathcal{C}$, we let $w, i \models \mathrm{P}_L\, \psi$ when there exists $j \in \mathsf{Pos}(w)$ such that $j < i$, $w(j,i) \in L$ and $w, j \models \psi$.

When no distinguished position is specified, it is customary to evaluate formulas at the *leftmost* unlabeled position. One could also consider the symmetrical convention of evaluating formulas at the *rightmost* unlabeled position. The convention chosen does not matter: we end-up with the same class of languages. However, we shall consider restrictions of $\mathrm{TL}[\mathcal{C}]$ for which this choice *does* matter. This is why we introduce notations for both conventions. Given a formula $\varphi \in \mathrm{TL}[\mathcal{C}]$ we let $L_{min}(\varphi) = \{w \in A^* \mid w, 0 \models \varphi\}$ and $L_{max}(\varphi) = \{w \in A^* \mid w, |w|+1 \models \varphi\}$.

We are now ready to define the operator $\mathcal{C} \mapsto \mathrm{TL}(\mathcal{C})$. Consider an arbitrary class $\mathcal{C}$. We write $\mathrm{TL}(\mathcal{C})$ for the class consisting of all languages $L_{min}(\varphi)$ where $\varphi \in \mathrm{TL}[\mathcal{C}]$. Observe that by definition, $\mathrm{TL}(\mathcal{C})$ is a Boolean algebra. Actually, the results of [27] imply that when $\mathcal{C}$ is a prevariety, then so is $\mathrm{TL}(\mathcal{C})$ (we do not need this fact in the present paper).

Classic unary temporal logic. Let $\text{ST} = \{\emptyset, A^*\}$ and $\text{DD} = \{\emptyset, \{\varepsilon\}, A^+, A^*\}$. The modalities F_{A^*} and P_{A^*} have the same semantics as the modalities F and P of standard unary temporal logic – *e.g.*, $w, i \models \text{F}\,\varphi$ when there exists $j \in \mathsf{Pos}(w)$ such that $i < j$ and $w, j \models \varphi$. Similarly, the modalities $\text{F}_{\{\varepsilon\}}$ and $\text{P}_{\{\varepsilon\}}$ have the same semantics as the modalities X (next) and Y (yesterday) – *e.g.*, $w, i \models \text{X}\,\varphi$ when $i{+}1 \in \mathsf{Pos}(w)$ and $w, i{+}1 \models \varphi$. Using these facts, one can check that the classes TL(ST) and TL(DD) correspond exactly to the two original standard variants of unary temporal logic (see *e.g.*, [9]): we have $\text{TL} = \text{TL(ST)}$ and $\text{TLX} = \text{TL(DD)}$.

▶ Remark 7 (Robustness of classes to which TL is applied). Note that including $\emptyset$ in an input class does not bring any new modality in unary temporal logic. Similarly, the classes TL(DD) and $\text{TL}(\text{DD} \setminus \{A^+\})$ are identical. However, in order to use generic results such as those from Section 3, we require the classes to which the operator $\mathcal{C} \mapsto \text{TL}(\mathcal{C})$ is applied to have robust properties: they should be prevarieties (hence, they should be closed under complement).

▶ Remark 8 (Connection with FO^2). Etessami, Vardi and Wilke [9] have shown that the variant TL corresponds to the class $\text{FO}^2(<)$ (two-variable first-order logic equipped with the linear order), and that TLX corresponds to $\text{FO}^2(<, +1)$ (which also allows the successor). In [27], these results are generalized to all classes $\text{TL}(\mathcal{C})$ where $\mathcal{C}$ is a *Boolean algebra*. In this case, we can construct from $\mathcal{C}$ a set of predicates $\mathbb{I}_{\mathcal{C}}$ such that $\text{TL}(\mathcal{C}) = \text{FO}^2(\mathbb{I}_{\mathcal{C}})$.

▶ Remark 9. Another important input is the class AT of alphabet testable languages. It consists of all Boolean combinations of languages A^*aA^*, where $a \in A$ is a letter. The class TL(AT) has been studied by Krebs, Lodaya, Pandya and Straubing [13, 14, 15], who worked with the definition based on two-variable first-order logic (*i.e.*, with the class $\text{FO}^2(\mathbb{I}_{\text{AT}})$, see Remark 8). In particular, they proved that TL(AT) has decidable membership. We shall obtain this result as a corollary of our generic characterization of the classes $\text{TL}(\mathcal{C})$.

4.2 Connection with unambiguous polynomial closure

It is shown in [27] that $\mathcal{C} \mapsto \text{TL}(\mathcal{C})$ can be expressed by other operators for very specific inputs: prevarieties of *group languages*. If $\mathcal{G}$ is such a class, then $\text{TL}(\mathcal{G})$ coincides with $\text{UPol}(\text{BPol}(\mathcal{G}))$, a class built on top of $\mathcal{G}$ with the two standard operators UPol and BPol. We do not use this result here, since we are tackling *arbitrary* input prevarieties, and in general, $\text{UPol}(\text{BPol}(\mathcal{C}))$ is *strictly* included in $\text{TL}(\mathcal{C})$ (it follows from [15] that the inclusion is strict for the class AT of Remark 9). However, the operators UPol and BPol remain key tools in the paper: we use two results of [27] about them. Let us first briefly recall their definitions.

Given finitely many languages $L_0, \dots, L_n \subseteq A^*$, a *marked product* of $L_0, \dots, L_n$ is a product of the form $L_0 a_1 L_1 \cdots a_n L_n$ where $a_1, \dots, a_n \in A$. A single language L_0 is a marked product (this is the case $n = 0$). The *polynomial closure* of a class $\mathcal{C}$, denoted by $\text{Pol}(\mathcal{C})$, consists of all *finite unions* of marked products $L_0 a_1 L_1 \cdots a_n L_n$ such that $L_0, \dots, L_n \in \mathcal{C}$. If $\mathcal{C}$ is a prevariety, then $\text{Pol}(\mathcal{C})$ is a lattice (this is due to Arfi [1], see also [19, 23] for recent proofs). However, $\text{Pol}(\mathcal{C})$ need not be closed under complement. This is why it is often combined with another operator: the Boolean closure of a class $\mathcal{D}$, denoted by $\text{Bool}(\mathcal{D})$, is the least Boolean algebra containing $\mathcal{D}$. We write $\text{BPol}(\mathcal{C})$ for $\text{Bool}(\text{Pol}(\mathcal{C}))$. It is standard that if $\mathcal{C}$ is a prevariety, then so is $\text{BPol}(\mathcal{C})$ (see [23] for example). Finally, UPol is the *unambiguous* restriction of Pol. A marked product $L_0 a_1 L_1 \cdots a_n L_n$ is *unambiguous* when every word $w \in L_0 a_1 L_1 \cdots a_n L_n$ has a *unique* decomposition $w = w_0 a_1 w_1 \cdots a_n w_n$ where $w_i \in L_i$ for $0 \leq i \leq n$. The *unambiguous polynomial closure* of a class $\mathcal{C}$, written $\text{UPol}(\mathcal{C})$, consists of all *finite disjoint unions* of *unambiguous marked products* $L_0 a_1 L_1 \cdots a_n L_n$ such that $L_0, \dots, L_n \in \mathcal{C}$ (by "disjoint" we mean that the languages in the union must be pairwise disjoint). While this is not apparent on the definition, it is known [27] that if the input class $\mathcal{C}$ is a prevariety, then so is $\text{UPol}(\mathcal{C})$. Thus, UPol preserves closure under complement.

In the paper, we are interested in the "combined" operator $\mathcal{C} \mapsto \text{UPol}(\text{BPol}(\mathcal{C}))$. Indeed, it is connected to the classes $\text{TL}(\mathcal{C})$ by the following proposition proved in [27, Proposition 9.12].

▶ **Proposition 10.** *For every prevariety $\mathcal{C}$, we have $\text{UPol}(\text{BPol}(\mathcal{C})) \subseteq \text{TL}(\mathcal{C})$.*

Although the inclusion of Proposition 10 is *strict* in general, it is essential for proving that *particular* languages belong to $\text{TL}(\mathcal{C})$. Indeed, we will combine it with the next result [27, Theorem 6.7] to prove that languages belong to $\text{UPol}(\text{BPol}(\mathcal{C}))$ – and therefore to $\text{TL}(\mathcal{C})$.

▶ **Theorem 11.** *Let $\mathcal{C}$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \to M$ be its syntactic morphism. Then, $L \in \text{UPol}(\text{BPol}(\mathcal{C}))$ if and only if α satisfies the following property:*

$$(esete)^{\omega+1} = (esete)^{\omega} ete(esete)^{\omega} \quad \textit{for every } \mathcal{C}\textit{-pair } (e,s) \in M^2 \textit{ and every } t \in M. \tag{1}$$

5 Algebraic characterization of TL($\mathcal{C}$)

We present a generic characterization of $\text{TL}(\mathcal{C})$ when $\mathcal{C}$ is a prevariety. It elegantly generalizes the characterizations of $\text{TL} = \text{FO}^2(<)$ and $\text{TLX} = \text{FO}^2(<, +1)$ by Thérien and Wilke [36] and that of $\text{TL}(\text{AT}) = \text{FO}^2(\mathbb{I}_{\text{AT}})$ by Krebs, Lodaya, Pandya and Straubing [13, 14, 15].

5.1 Statement

The characterization is based on the well-known variety of finite monoids DA (see [35] for a survey on this class). A finite monoid M belongs to DA if it satisfies the following equation:

$$(st)^{\omega} = (st)^{\omega} t (st)^{\omega} \quad \text{for every } s, t \in M. \tag{2}$$

Thérien and Wilke [36] showed that a regular language belongs to TL if and only if its syntactic monoid is in DA (strictly speaking, they considered two-variable first-order logic, the equality $\text{FO}^2(<) = \text{TL}$ is due to Etessami, Vardi and Wilke [9]). We extend this result in the following generic characterization of $\text{TL}(\mathcal{C})$, based on $\mathcal{C}$-orbits introduced in Section 3.

▶ **Theorem 12.** *Let $\mathcal{C}$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \to M$ be its syntactic morphism. The two following properties are equivalent:*

1. *$L \in \text{TL}(\mathcal{C})$.*
2. *For every idempotent $e \in E(M)$, the $\mathcal{C}$-orbit of e for α belongs to* DA.

Given as input a regular language $L \subseteq A^*$, one can compute its syntactic morphism $\alpha : A^* \to M$. In view of Theorem 12, $L \in \text{TL}(\mathcal{C})$ if and only if for every $e \in E(M)$, the $\mathcal{C}$-orbit of e for α belongs to DA. The latter condition can be decided by checking all $\mathcal{C}$-orbits, provided that we are able to compute them. By Lemma 5, this is possible when $\mathcal{C}$-separation is decidable. Altogether, we obtain the following corollary of Theorem 12.

▶ **Corollary 13.** *If a prevariety $\mathcal{C}$ has decidable separation, $\text{TL}(\mathcal{C})$ has decidable membership.*

▶ Remark 14. Let $L \subseteq A^*$ be a regular language and $\alpha : A^* \to M$ be its syntactic morphism. The fact that the $\mathcal{C}$-orbit of $e \in E(M)$ for α belongs to DA means that we have,

$$(esete)^{\omega} = (esete)^{\omega} ete(esete)^{\omega} \quad \text{for all } s, t \in M \text{ such that } (e,s) \text{ and } (e,t) \text{ are } \mathcal{C}\text{-pairs.} \tag{3}$$

One can check that (3) follows from (1), which characterizes $\text{UPol}(\text{BPol}(\mathcal{C}))$ (this is consistent with Proposition 10 asserting that $\text{UPol}(\text{BPol}(\mathcal{C})) \subseteq \text{TL}(\mathcal{C})$). Indeed, choosing $t = s$ in (1) shows that the $\mathcal{C}$-orbit of e is aperiodic, *i.e.*, $(ese)^{\omega+1} = (ese)^{\omega}$ if (e, s) is a $\mathcal{C}$-pair. However, note that the element t is "free" in (1), whereas it must be part of a $\mathcal{C}$-pair (e, t) in (3).

Before proving Theorem 12, we first explain why it generalizes the original characterizations of the classes TL, TLX and TL(AT), as mentioned at the beginning of the section.

5.2 Application to historical classes

We first deduce the original characterizations of the classes $\mathrm{TL} = \mathrm{TL}(\mathrm{ST})$ and $\mathrm{TLX} = \mathrm{TL}(\mathrm{DD})$ by Thérien and Wilke [36] as simple corollaries of Theorem 12. We start with the former.

▶ **Theorem 15** (Thérien and Wilke [36]). *Let $L \subseteq A^*$ be a regular language and let M be its syntactic monoid. The two following properties are equivalent:*
1. *L belongs to* TL.
2. *M belongs to* DA.

Proof. Let $\alpha : A^* \to M$ be the syntactic morphism of L. Since $\mathrm{TL} = \mathrm{TL}(\mathrm{ST})$, Theorem 12 implies that $L \in \mathrm{TL}$ if and only if every ST-orbit for α belongs to DA. Since $\mathrm{ST} = \{\emptyset, A^*\}$, every pair $(e, s) \in E(M) \times M$ is a $\mathcal{C}$-pair, so that the ST-orbit of $e \in E(M)$ for α is eMe. In particular the ST-orbit of 1_M is the whole monoid M. Hence, every ST-orbit for α belongs to DA if and only if M belongs to DA, which completes the proof. ◀

We turn to the characterization of $\mathrm{TLX} = \mathrm{TL}(\mathrm{DD})$, also due to Thérien and Wilke [36]. In order to state it, we need an additional definition. Consider a regular language L and let $\alpha : A^* \to M$ be its syntactic morphism. The *syntactic semigroup of L* is the set $S = \alpha(A^+)$. Note that for every idempotent $e \in E(S)$, the set eSe is a monoid whose neutral element is e.

▶ **Theorem 16** (Thérien and Wilke [36]). *Let $L \subseteq A^*$ be a regular language and S be its syntactic semigroup. The two following properties are equivalent:*
1. *L belongs to* TLX.
2. *For every $e \in E(S)$, the monoid eSe belongs to* DA.

Proof. Let $\alpha : A^* \to M$ be the syntactic morphism of L. For $e \in E(M)$, let $M_e \subseteq M$ be the DD-orbit of e for α. Since $\mathrm{DD} = \{\emptyset, \{\varepsilon\}, A^+, A^*\}$, for $(e, s) \in E(S) \times S$, the language $\alpha^{-1}(e)$ is not DD-separable from $\alpha^{-1}(s)$. Hence, (e, s) is a $\mathcal{C}$-pair, so that $M_e = eSe$ for all $e \in E(S)$. Moreover, if $1_M \notin E(S)$ (which means that $\alpha^{-1}(1_M) = \{\varepsilon\}$), then we have $M_{1_M} = \{1_M\}$ (which clearly belongs to DA). Hence, every DD-orbit for α belongs to DA if and only if $eSe \in \mathsf{DA}$ for every $e \in E(S)$. In view of Theorem 12, this implies Theorem 16. ◀

Finally, we consider the class TL(AT), defined and characterized by Krebs, Lodaya, Pandya and Straubing [13, 14, 15]. Let us first present their characterization. It is based on a variety of finite monoids called $\mathsf{M}_e\mathsf{DA}$. Let M be a finite monoid. For each $e \in E(M)$, let $N_e \subseteq M$ be the submonoid of M generated by the set $\{s \in M \mid e \leqslant_{\mathscr{J}} s\}$. We say that M belongs to $\mathsf{M}_e\mathsf{DA}$ if and only if for every idempotent $e \in E(M)$, the monoid of eN_ee belongs to DA.

▶ **Theorem 17** (Krebs, Lodaya, Pandya and Straubing [15]). *Let $L \subseteq A^*$ be a regular language and M be its syntactic monoid. The two following properties are equivalent:*
1. *$L \in \mathrm{TL}(\mathrm{AT})$.*
2. *M belongs to $\mathsf{M}_e\mathsf{DA}$.*

Proof. For $w \in A^*$, let $alph(w) \subseteq A$ be the set of letters occurring in w (*i.e.*, the least set $B \subseteq A$ such that $w \in B^*$). For $e \in E(M)$, let M_e be the AT-orbit of e for α. We prove that $M_e = eN_ee$ for every $e \in E(M)$. It will follows that M belongs to $\mathsf{M}_e\mathsf{DA}$ if and only if every AT-orbit for α belongs to DA. In view of Theorem 12 this implies Theorem 17.

We first consider $s' \in eN_ee$ and prove that $s' \in M_e$. We have $s \in N_e$ such that $s' = ese$. By definition, $s = s_1 \cdots s_n$ where $e \leqslant_{\mathscr{J}} s_i$ for every $i \leq n$. If $n = 0$, then $s = 1_M$ and $ese = e \in M_e$. Assume now that we have $n \geq 1$. Since $e \leqslant_{\mathscr{J}} s_i$, we have $q_i, r_i \in M$ such that $e = q_is_ir_i$ for every $i \leq n$. Hence, since $e \in E(M)$, we have $e = q_1s_1r_1 \cdots q_ns_nr_n$. For every

$i \leq n$, let $x_i \in \alpha^{-1}(q_i)$, $y_i \in \alpha^{-1}(r_i)$ and $u_i \in \alpha^{-1}(s_i)$. Finally, let $w = x_1u_1y_1 \cdots x_nu_ny_n$ and $w' = wu_1 \cdots u_nw$. By definition, we have $e = \alpha(w)$ and $ese = \alpha(w')$. Moreover, it is clear that $alph(w) = alph(w')$. By definition of AT, it follows that $\alpha^{-1}(e)$ is not AT-separable from $\alpha^{-1}(ese)$. Thus, (e, ese) is an AT-pair for α, which yields $s' = ese \in M_e$, as desired.

Conversely, let $s' \in M_e$. By definition, there exists an AT-pair $(e,s) \in M^2$ with $e \in E(M)$ such that $s' = ese$. Therefore, by definition of AT, there exist $u, v \in A^*$ such that $alph(u) = alph(v)$, $\alpha(u) = e$ and $\alpha(v) = s$. Let $a_1 \dots, a_n \in A$ be the letters such that $v = a_1 \cdots a_n$. Since $alph(u) = alph(v)$, it is immediate that for each $i \leq n$, there are $x_i, y_i \in A^*$ such that $u = x_ia_iy_i$. Hence $e = \alpha(u) \leqslant_{\mathscr{J}} \alpha(a_i)$ and we conclude that $s = \alpha(a_1 \cdots a_n) \in N_e$. Consequently, $s' = ese \in eN_ee$, as desired. ◀

5.3 Proof of Theorem 12

We fix a prevariety $\mathcal{C}$, a regular language $L \subseteq A^*$ and its syntactic morphism $\alpha : A^* \to M$ for the proof. We prove that $L \in \mathrm{TL}(\mathcal{C})$ if and only if all $\mathcal{C}$-orbits for α belong to DA. We start with the left-to-right implication.

From TL($\mathcal{C}$) to DA. This direction follows from results of [27]. To use them, we need some preliminary terminology. We introduce equivalence relations connected to the class $\mathrm{TL}(\mathcal{C})$ when $\mathcal{C}$ is a prevariety. Given a morphism $\eta : A^* \to N$ into a finite monoid N, denote by $\mathcal{C}_\eta$ be the class of all languages recognized by η. The following fact is easy (see [27, Fact 9.3]).

▶ **Fact 18.** *Let $\mathcal{C}$ be a prevariety. For every $\mathrm{TL}[\mathcal{C}]$ formula φ, there exists a $\mathcal{C}$-morphism $\eta : A^* \to N$ such that φ is a $\mathrm{TL}[\mathcal{C}_\eta]$ formula.*

We use the standard notion of rank of a $\mathrm{TL}[\mathcal{C}_\eta]$ formula: the *rank* of φ is defined as the length of the longest sequence of nested temporal operators within its parse tree. Formally:
- Any atomic formula has rank 0.
- The rank of $\neg\varphi$ is the same as the rank of φ.
- The rank of $\varphi \vee \psi$ and $\varphi \wedge \psi$ is the maximum between the ranks of φ and ψ.
- For every language $L \subseteq A^*$, the rank of $\mathrm{F}_L\, \varphi$ and $\mathrm{P}_L\, \varphi$ is the rank of φ plus 1.

Two $\mathrm{TL}[\mathcal{C}_\eta]$ formulas φ and ψ are *equivalent* if they have the same semantics. That is, for every $w \in A^*$ and every position $i \in \mathsf{Pos}(w)$, 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 19.** *Let $\eta : A^* \to N$ be a morphism into a finite monoid and let $k \in \mathbb{N}$. There are only finitely many non-equivalent $\mathrm{TL}[\mathcal{C}_\eta]$ formulas of rank at most k.*

We now define equivalence relations. Let $\eta : A^* \to N$ be a morphism into a finite monoid and let $k \in \mathbb{N}$. Given $w, w' \in A^*$, $i \in \mathsf{Pos}(w)$ and $i' \in \mathsf{Pos}(w')$, we write, $w, i \cong_{\eta,k} w', i'$ when:

$$\text{For every } \mathrm{TL}[\mathcal{C}_\eta] \text{ formula } \varphi \text{ of rank at most } k, \quad w, i \models \varphi \Longleftrightarrow w', i' \models \varphi.$$

It is straightforward that $\cong_{\eta,k}$ is an equivalence relation. Moreover, it is immediate from the definition and Lemma 19, that $\cong_{\eta,k}$ has finite index. We lift each relation $\cong_{\eta,k}$ to A^* (abusing terminology, we also denote by $\cong_{\eta,k}$ the new relation): given $w, w' \in A^*$, we write $w \cong_{\eta,k} w'$ when $w, 0 \cong_{\eta,k} w', 0$. Clearly, $\cong_{\eta,k}$ is an equivalence relation of finite index over A^*. Moreover, we have the following connection between $\mathrm{TL}(\mathcal{C})$ and the relations $\cong_{\eta,k}$.

▶ **Lemma 20.** *Let $\mathcal{C}$ be a prevariety and $L \subseteq A^*$. If $L \in \mathrm{TL}(\mathcal{C})$, then there exists a $\mathcal{C}$-morphism $\eta : A^* \to N$ and $k \in \mathbb{N}$ such that L is a union of $\cong_{\eta,k}$-classes.*

Proof. Let $L \in \mathrm{TL}(\mathcal{C})$. There exists a $\mathrm{TL}[\mathcal{C}]$ formula φ such that $w \in L \Leftrightarrow w, 0 \models \varphi$ for all $w \in A^*$. By Fact 18, there exists a $\mathcal{C}$-morphism $\eta : A^* \to N$ such that φ is a $\mathrm{TL}[\mathcal{C}_\eta]$ formula. Let $k \in \mathbb{N}$ be the rank of φ. We prove that L is a union of $\cong_{\eta,k}$-classes. Given $w, w' \in A^*$ such that $w \cong_{\eta,k} w'$, we have to prove that $w \in L \Leftrightarrow w' \in L$. By symmetry, we only prove the left to right implication. Thus, we assume that $w \in L$. By definition of φ, it follows that $w, 0 \models \varphi$. Moreover, since $w \cong_{\eta,k} w'$ (*i.e.*, $w, 0 \cong_{\eta,k} w', 0$) and φ is a $\mathrm{TL}[\mathcal{C}_\eta]$ formula of rank k, we have $w', 0 \models \varphi$ by definition of $\cong_{\eta,k}$. Hence, $w' \in L$ by definition of φ, as desired. ◀

In addition to the link stated in Lemma 20 between $\mathrm{TL}(\mathcal{C})$ and the equivalence relations $\cong_{\eta,k}$, we use a property of $\cong_{\eta,k}$ that follows from [27, Lemma 9.6 and Proposition 9.7].

▶ **Proposition 21.** *Consider a morphism $\eta : A^* \to N$ into a finite monoid, let $f \in E(N)$ be an idempotent, let $u, v, z \in \eta^{-1}(f)$ and let $x, y \in A^*$. For every $k \in \mathbb{N}$, we have:*

$$x(z^k uz^{2k} vz^k)^k (z^k uz^{2k} vz^k)^k y \cong_{\eta,k} x(z^k uz^{2k} vz^k)^k z^k vz^k (z^k uz^{2k} vz^k)^k y.$$

We are ready to conclude this direction of the proof: assuming that $L \in \mathrm{TL}(\mathcal{C})$, we show that all $\mathcal{C}$-orbits for its syntactic monoid belong to DA. Let $e \in E(M)$ and M_e be its $\mathcal{C}$-orbit. Proving that $M_e \in \mathsf{DA}$ amounts to proving that any elements $s, t \in M_e$ satisfy (2). Fix $e, s, t \in E(M) \times M_e \times M_e$. Lemma 20 yields a $\mathcal{C}$-morphism $\eta : A^* \to N$ and $k \in \mathbb{N}$ such that L is a union of $\cong_{\eta,k}$-classes. Since η is a $\mathcal{C}$-morphism, Lemma 6 yields $f \in E(N)$ such that $M_e \subseteq \alpha(\eta^{-1}(f))$. Since $e, s, t \in M_e$, we get $z, u, v \in A^*$ such that $z, u, v \in \eta^{-1}(f)$, $\alpha(z) = e$, $\alpha(u) = s$ and $\alpha(v) = t$. Let $x, y \in A^*$ be two arbitrary words. By Proposition 21, we obtain,

$$x(z^k uz^{2k} vz^k)^k (z^k uz^{2k} vz^k)^k y \cong_{\eta,k} x(z^k uz^{2k} vz^k)^k z^k vz^k (z^k uz^{2k} vz^k)^k y.$$

Since L is a union of $\cong_{\eta,k}$-classes, the words $(z^k uz^k vz^k)^{2k}$ and $(z^k uz^k vz^k)^k z^k vz^k (z^k uz^k vz^k)^k$ are equivalent for the syntactic congruence of L, so they have the same image under its syntactic morphism α. Since $e \in E(M)$, this yields $(esete)^{2k} = (esete)^k ete (esete)^k$. Hence, $(st)^{2k} = (st)^k t (st)^k$ since $e, s, t \in M_e$ and e is neutral in M_e by Lemma 4. It now suffices to multiply by enough copies of st on both sides to get $(st)^\omega = (st)^\omega t (st)^\omega$. Therefore, (2) holds.

From DA to TL($\mathcal{C}$). Assuming that every $\mathcal{C}$-orbit for the syntactic morphism $\alpha : A^* \to M$ of L belongs to DA, we have to show that $L \in \mathrm{TL}(\mathcal{C})$, *i.e.*, to build a $\mathrm{TL}[\mathcal{C}]$ formula defining L. Let us start by giving a high-level overview of the proof for this direction.

Since $\mathrm{TL}(\mathcal{C})$ is closed under union, it suffices to prove that for all $s \in M$, the language $\alpha^{-1}(s)$ is in $\mathrm{TL}(\mathcal{C})$. We achieve this by inductively constructing a $\mathrm{TL}[\mathcal{C}]$ formula defining $\alpha^{-1}(s)$. According to Lemma 3, there exists a $\mathcal{C}$-morphism $\eta : A^* \to N$ such that the $\mathcal{C}$-pairs for α are exactly the η-pairs for α. We use η to leverage the assumption that all $\mathcal{C}$-orbits for α belong to DA. More precisely, η recognizes all the basic languages in $\mathcal{C}$ that we shall use in our $\mathrm{TL}[\mathcal{C}]$ formulas. The induction proceeds as follows: using η, we define a sequence of languages $K_0 \supseteq K_1 \supseteq \cdots \supseteq K_{|N|}$ and show by induction on $|N| - \ell$ that $K_\ell \cap \alpha^{-1}(s)$ can be defined by a $\mathrm{TL}[\mathcal{C}]$ formula for each $\ell \leq |N|$. The induction basis is the case $\ell = |N|$, which is simple because $K_{|N|}$ is a finite language. Furthermore, the case $\ell = 0$ gives the desired result since K_0 contains all words. The induction step consists in building a $\mathrm{TL}[\mathcal{C}]$ formula describing $K_\ell \cap \alpha^{-1}(s)$ from several $\mathrm{TL}[\mathcal{C}]$ formulas that describe the languages $K_{\ell+1} \cap \alpha^{-1}(t)$ for all $t \in M$. However, the actual argument is slightly more involved. Indeed, in order to perform the induction step, we must abstract each word in $K_\ell \cap \alpha^{-1}(s)$ by considering a specific decomposition of this word and viewing each infix as a new letter. We then argue that the resulting word belongs to $K_{\ell+1} \cap \alpha^{-1}(s)$, which allows us to apply induction. Yet, for this process to work, the letter that we use to abstract an infix must have the same images as this original infix under both α and η. This is problematic, because such a letter does not necessarily exist. We solve this issue by considering an extended alphabet B, replacing $\alpha : A^* \to M$ and $\eta : A^* \to N$ with two new morphisms $\beta : B^* \to M$ and $\delta : B^* \to N$ that have the required property. Of course, this involves some preliminary work: we must reformulate both our objective (proving that all languages $\alpha^{-1}(s)$ can be defined in $\mathrm{TL}(\mathcal{C})$) and our hypothesis (that every $\mathcal{C}$-orbit for α belongs to DA) on the new morphisms β and η.

We now start the proof by first defining β and δ. Recall that $\eta : A^* \to N$ is the $\mathcal{C}$-morphism provided by Lemma 3: it is such that the $\mathcal{C}$-pairs for α are exactly the η-pairs for α. We fix η for the entire proof. We define an auxiliary alphabet B. Let $P \subseteq M \times N$ be the set of all pairs $(\alpha(w), \eta(w)) \in M \times N$ where $w \in A^+$ is a nonempty word. For each pair $(s, r) \in P$, we create a fresh letter $b_{s,r} \notin A$ and we define $B = \{b_{s,r} \mid (s, r) \in P\}$.

Let $\beta : B^* \to M$ and $\delta : B^* \to N$ be the morphisms defined by $\beta(b_{s,r}) = s$ and $\delta(b_{s,r}) = r$ for $(s, r) \in P$. By definition, we have $(\beta(w), \delta(w)) \in P$ for all $w \in B^+$. Let $\mathcal{C}_\delta$ be the class of all languages over B recognized by δ. One can check that $\mathcal{C}_\delta$ is a prevariety. We now reduce membership of inverse images under α to $\mathrm{TL}(\mathcal{C})$ to that of inverse images under β to $\mathrm{TL}(\mathcal{C}_\delta)$.

▶ **Lemma 22.** *For every $F \subseteq M$, if $\beta^{-1}(F) \in \mathrm{TL}(\mathcal{C}_\delta)$, then $\alpha^{-1}(F) \in \mathrm{TL}(\mathcal{C})$.*

Proof. We first define a morphism $\gamma : A^* \to B^*$. Consider a letter $a \in A$. By definition, $(\alpha(a), \eta(a)) \in P$. Hence, we may define $\gamma(a) = b_{\alpha(a),\eta(a)} \in B$. By definition, we have $\alpha(w) = \beta(\gamma(w)) \in M$ and $\eta(w) = \delta(\gamma(w))$ for every $w \in A^*$. It follows that for every $F \subseteq M$, we have $\alpha^{-1}(F) = \gamma^{-1}(\beta^{-1}(F)) \subseteq A^*$. Consequently, it now suffices to prove that for every $K \subseteq B^*$ such that $K \in \mathrm{TL}(\mathcal{C}_\delta)$, we have $\gamma^{-1}(K) \in \mathrm{TL}(\mathcal{C})$. We fix K for the proof. Since $K \in \mathrm{TL}(\mathcal{C}_\delta)$, it is defined by a formula $\psi \in \mathrm{TL}[\mathcal{C}_\delta]$. We apply two kinds of modifications to ψ in order to build a new formula $\psi' \in \mathrm{TL}[\mathcal{C}]$ defining $\gamma^{-1}(K)$:

1. We replace every atomic subformula "b" for $b \in B$ by the $\mathrm{TL}[\mathcal{C}]$-formula $\bigvee_{\{a \in A \mid \gamma(a) = b\}} a$.
2. For every temporal modality F_H (resp. P_H) occurring in ψ, we have $H \in \mathcal{C}_\delta$ by hypothesis. Hence, H is recognized by δ and there exists $G \subseteq N$ such that $H = \delta^{-1}(G)$. Note that $\eta^{-1}(G) \in \mathcal{C}$ since η is a $\mathcal{C}$-morphism. We replace the temporal modality F_H (resp. P_H) by $\mathrm{F}_{\eta^{-1}(G)}$ (resp. $\mathrm{P}_{\eta^{-1}(G)}$).

By definition the resulting formula ψ' belongs to $\mathrm{TL}[\mathcal{C}]$ and one can verify that for every $w \in A^*$, we have $w, 0 \models \psi' \Leftrightarrow \gamma(w), 0 \models \psi$. Since $L_{min}(\psi) = K$, we get $L_{min}(\psi') = \gamma^{-1}(K)$, which implies that $K \in \mathrm{TL}(\mathcal{C})$. This completes the proof. ◀

In view of Lemma 22, it suffices to prove that any language recognized by β belongs to $\mathrm{TL}(\mathcal{C}_\delta)$. Since L is recognized by α, this will imply $L \in \mathrm{TL}(\mathcal{C})$, which is our goal. In the next lemma, we reformulate on β and δ the assumption that every $\mathcal{C}$-orbit for α belongs to DA.

▶ **Lemma 23.** *For every $e \in E(M)$ and every $s, t \in M$, if (e, s) and (e, t) are δ-pairs for β, then $(esete)^\omega = (esete)^\omega ete(esete)^\omega$.*

Proof. By hypothesis, there exist $u, v, x, y \in B^*$ such that $\delta(u) = \delta(v)$, $\delta(x) = \delta(y)$, $\beta(u) = \beta(x) = e$, $\beta(v) = s$ and $\beta(y) = t$. The definitions of β and δ imply that for any $w \in B^*$, there exists $w' \in A^*$ such that $\delta(w) = \eta(w')$ and $\beta(w) = \alpha(w')$. Therefore, we obtain $u', v', x', y' \in A^*$ such that $\eta(u') = \eta(v')$, $\eta(x') = \eta(y')$, $\alpha(u') = \alpha(x') = e$, $\alpha(v') = s$ and $\alpha(y') = t$. Thus, $(e, s) \in M^2$ and $(e, t) \in M^2$ are η-pairs for α. By definition of η, it follows that they are $\mathcal{C}$-pairs for α. Hence, ese and ete both belong to the $\mathcal{C}$-orbit of e for α. Since all $\mathcal{C}$-orbits for α belong to DA by hypothesis, this gives $(esete)^\omega = (esete)^\omega ete(esete)^\omega$. ◀

We now use the Green relation $\mathcal{J}$ over N to associate a number $d_{\mathcal{J}}(r) \in \mathbb{N}$ with every element $r \in N$. We let $d_{\mathcal{J}}(r)$ be the maximal number $n \in \mathbb{N}$ such that there exist n elements $r_1, \dots, r_n \in N$ satisfying $r <_{\mathcal{J}} r_1 <_{\mathcal{J}} \cdots <_{\mathcal{J}} r_n$. By definition, $0 \leq d_{\mathcal{J}}(r) \leq |N| - 1$. In particular, we have $d_{\mathcal{J}}(r) = 0$ if and only if r is maximal for $\leqslant_{\mathcal{J}}$ (*i.e.*, if and only if $r \mathrel{\mathcal{J}} 1_N$). Finally, given a word $w \in B^*$, we write $d_{\mathcal{J}}(w) \in \mathbb{N}$ for $d_{\mathcal{J}}(\delta(w))$. Observe that for all $x, y, z \in B^*$, we have $d_{\mathcal{J}}(y) \leq d_{\mathcal{J}}(xyz)$ (as $xyz \leqslant_{\mathcal{J}} y$), a fact that we shall use frequently.

In order to argue inductively, we define a family of languages $K_\ell \subseteq B^*$ for $\ell \in \mathbb{N}$ as follows:

$$K_\ell = \left\{ w \in B^* \mid \text{for all } k \leq \ell \text{ and } x, y, z \in B^*, \text{ if } w = xyz \text{ and } |y| = k, \text{ then } d_{\mathcal{J}}(y) \geq k \right\}.$$

Note that $K_0 = B^*$ as $d_{\mathcal{J}}(y) \geq 0$ for all $y \in B^*$. Also, if $\ell \geq |N|$, then K_ℓ is *finite* (it contains words of length at most $|N| - 1$ as $d_{\mathcal{J}}(y) < |N|$ for all $y \in B^*$). We now have the next lemma.

▶ **Lemma 24.** *Let $\ell \in \mathbb{N}$ and $w \in K_\ell$. Then $d_{\mathcal{J}}(w) \leq \ell$ if and only if for all $x, y, z \in B^*$ such that $w = xyz$ and $|y| \leq \ell + 1$, we have $d_{\mathcal{J}}(y) \leq \ell$.*

Proof. The "only if" direction is immediate since $d_{\mathcal{J}}(y) \leq d_{\mathcal{J}}(w)$ for every infix y of w. Conversely, assume that for all $x, y, z \in B^*$ such that $w = xyz$ and $|y| \leq \ell + 1$, we have $d_{\mathcal{J}}(y) \leq \ell$. We prove that $d_{\mathcal{J}}(w) \leq \ell$. If $|w| \leq \ell + 1$, this is immediate. Assume now that $|w| > \ell + 1$. We get $b_1, \dots, b_n \in B$ and $v \in B^*$ such that $|v| = \ell + 1$ and $w = vb_1 \cdots b_n$. We use induction on i to prove that $\delta(v) \mathrel{\mathcal{J}} \delta(vb_1 \cdots b_i)$ for all $i \leq n$. Since $d_{\mathcal{J}}(v) \leq \ell$ by hypothesis, the case $i = n$ yields $d_{\mathcal{J}}(w) \leq \ell$. The case $i = 0$ is trivial: we have $\delta(v) \mathrel{\mathcal{J}} \delta(v)$. Assume now that $i \geq 1$. By induction hypothesis, we know that $\delta(v) \mathrel{\mathcal{J}} \delta(vb_1 \cdots b_{i-1})$. Let $x, y \in B^*$ such that $|y| = \ell$ and $xy = vb_1 \cdots b_{i-1}$ (the words x and y exist because $|v| = \ell + 1$). Since $w \in K_\ell$, and y is an infix of w such that $|y| = \ell$, we know that $d_{\mathcal{J}}(y) \geq \ell$. Moreover, yb_i is an infix of w such that $|yb_i| = \ell + 1$, which yields $d_{\mathcal{J}}(yb_i) \leq \ell$ by hypothesis. Since $d_{\mathcal{J}}(y) \leq d_{\mathcal{J}}(yb_i)$, we get $d_{\mathcal{J}}(yb_i) = d_{\mathcal{J}}(y) = \ell$, which implies that $\delta(yb_i) \mathrel{\mathcal{J}} \delta(y)$. Moreover, we have $\delta(yb_i) \leqslant_{\mathcal{R}} \delta(y)$. Thus, Lemma 1 yields $\delta(yb_i) \mathrel{\mathcal{R}} \delta(y)$. This implies that $\delta(xyb_i) \mathrel{\mathcal{R}} \delta(xy)$. Hence, $\delta(vb_1 \cdots b_i) \mathrel{\mathcal{J}} \delta(vb_1 \cdots b_{i-1}) \mathrel{\mathcal{J}} \delta(v)$. This completes the proof. ◀

We now prove that for all $s \in M$ and $\ell \in \mathbb{N}$, we have $K_\ell \cap \beta^{-1}(s) \in \mathrm{TL}(\mathcal{C}_\delta)$. Our objective (every language recognized by β belongs to $\mathrm{TL}(\mathcal{C}_\delta)$) follows from the case $\ell = 0$, since $K_0 = B^*$. The proof involves two steps. The first settles the case of elements of $K_\ell \cap \beta^{-1}(s)$ whose image under δ has a $d_{\mathcal{J}}$ value at most ℓ. We do not use induction for this case, which relies on the inclusion $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C}_\delta)) \subseteq \mathrm{TL}(\mathcal{C}_\delta)$. It is also the place where we use Lemma 23, *i.e.*, the hypothesis that all $\mathcal{C}$-orbits for α are in DA.

▶ **Proposition 25.** *Let $(\ell, s, r) \in \mathbb{N} \times M \times N$. If $d_{\mathcal{J}}(r) \leq \ell$ then $K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r) \in \mathrm{TL}(\mathcal{C}_\delta)$.*

Proof. We prove that $K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r) \in \mathrm{UPol}(\mathrm{BPol}(\mathcal{C}_\delta))$, which, by Proposition 10, will give the desired result $K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r) \in \mathrm{TL}(\mathcal{C}_\delta)$. Let $\gamma : B^* \to Q$ be the syntactic morphism of $K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r)$. By Theorem 11, it suffices to show that given $q_1, q_2 \in Q$ and $f \in E(Q)$ such that $(f, q_1) \in Q^2$ is a $\mathcal{C}_\delta$-pair for γ, the following equation holds:

$$(fq_1fq_2f)^{\omega+1} = (fq_1fq_2f)^{\omega} fq_2f (fq_1fq_2f)^{\omega}. \tag{4}$$

Let $q_1, q_2, f \in Q$ be such elements. By definition of $\mathcal{C}_\delta$, we know that δ is a $\mathcal{C}_\delta$-morphism. Therefore, Lemma 3 implies that (f, q_1) is a δ-pair for γ. We get $u', v'_1 \in B^*$, such that $\delta(u') = \delta(v'_1)$, $\gamma(u') = f$ and $\gamma(v'_1) = q_1$. Note that if $v'_1 = \varepsilon$, then $q_1 = 1_Q$ and (4) holds since it is clear that $(fq_2f)^{\omega+1} = (fq_2f)^{2\omega+1}$. Therefore, we assume from now on that $v'_1 \in B^+$. Let us also choose $v'_2 \in B^*$ such that $\gamma(v'_2) = q_2$. We now define $p = \ell \times \omega(N) \times \omega(M) \times \omega(Q)$,

$u = (u')^p$, $v_1 = (u')^{p-1}v'_1$ and $v_2 = uv'_2u(uv_1uv'_2u)^{p-1}$. We compute $\gamma(u) = f$, $\gamma(v_1) = fq_1$ and $\delta(u) = \delta(v_1)$. Moreover, since p is a multiple of $\omega(N)$, the element $\delta(u) = \delta(v_1)$ is an idempotent $g \in E(N)$. Finally, we have $\gamma(v_2) = fq_2f(fq_1fq_2f)^{p-1}$ and $\delta(v_2) = (g\delta(v'_2)g)^p$. In particular, it follows that $\delta(v_2)$ is an idempotent $h \in E(N)$ such that $gh = hg = h$.

We prove that $(uv_1uv_2u)^p$ and $(uv_1uv_2u)^p uv_2u(uv_1uv_2u)^p$ are equivalent for the syntactic congruence of $K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r)$. This will imply that they have the same image under γ, which yields $(fq_1fq_2f)^\omega = (fq_1fq_2f)^\omega fq_2f(fq_1fq_2f)^{2\omega-1}$. One may then multiply by fq_1fq_2f on the right to get (4), as desired. For $x, y \in A^*$, let $z_1 = x(uv_1uv_2u)^p y$ and $z_2 = x(uv_1uv_2u)^p uv_2u(uv_1uv_2u)^p y$. We have to show that $z_1 \in K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r)$ if and only if $z_2 \in K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r)$. We first treat the special case where $|u| < \ell$.

Assume that $|u| < \ell$. We show that in this case $z_1 \notin K_\ell$ and $z_2 \notin K_\ell$ (which implies the desired result). Since $u = (u')^p$ and $p \geq \ell$, the hypothesis that $|u| < \ell$ yields $u = u' = \varepsilon$. Since $\delta(u) = \delta(v_1)$, we get $\delta(v_1) = 1_N$. Recall that $v_1 = (u')^{p-1}v'_1$ and $v'_1 \in B^+$ by hypothesis. Thus, $v_1 \in B^+$, which means that it contains a letter $b \in B$ such that $\delta(b) \mathrel{\mathcal{J}} 1_N$. In particular $d_{\mathcal{J}}(b) = 0$. Hence, b is an infix of length 1 of both z_1 and z_2 such that $d_{\mathcal{J}}(b) < 1$. Now $\ell > |u| = 0$, so that $\ell \geq 1$. This implies $z_1 \notin K_\ell$ and $z_2 \notin K_\ell$. This completes the special case.

From now on, we assume that $|u| \geq \ell$. Since $\delta(u) = \delta(v_1) = g \in E(N)$, $\delta(v_2) = h \in E(N)$ and $gh = hg = h$, we have $\delta(z_1) = \delta(z_2) = \delta(x)h\delta(y)$. Therefore, $z_1 \in \delta^{-1}(r)$ if and only if $z_2 \in \delta^{-1}(r)$. Let us prove that $z_1 \in K_\ell \Leftrightarrow z_2 \in K_\ell$. This is trivial if $\ell = 0$ since $K_0 = B^*$. Assume now that $\ell \geq 1$. Since $|u| \geq \ell$ by hypothesis, it follows that for every $k \leq \ell$, z_1 and z_2 have the same infixes of length k. This implies that $z_1 \in K_\ell \Leftrightarrow z_2 \in K_\ell$, as desired.

It remains to prove that if $z_1, z_2 \in K_\ell \cap \delta^{-1}(r)$, then $\beta(z_1) = \beta(z_2)$. We first show that our assumptions imply $g \mathrel{\mathcal{J}} h$. Again, there are two cases. First, assume that $\ell = 0$. Since $d_{\mathcal{J}}(r) \leq \ell$ by hypothesis, we get $r \mathrel{\mathcal{J}} 1_N$. Thus, since u and v_2 are infixes of $z_1 \in \delta^{-1}(r)$, we have $\delta(u) \mathrel{\mathcal{J}} \delta(v_2) \mathrel{\mathcal{J}} 1_N$, which exactly says that $g \mathrel{\mathcal{J}} h \mathrel{\mathcal{J}} 1_N$. Assume now that $\ell \geq 1$. Recall that $|u| \geq \ell$. Since u is an infix of v_2, this also implies that $|v_2| \geq \ell$. Hence, since u and v_2 are infixes of $z_2 \in K_\ell \cap \delta^{-1}(r)$, we get $d_{\mathcal{J}}(u) \geq \ell$ and $d_{\mathcal{J}}(v_2) \geq \ell$, $r \leqslant_{\mathcal{J}} \delta(u)$ and $r \leqslant_{\mathcal{J}} \delta(v_2)$. In particular, it follows that $d_{\mathcal{J}}(r) \geq d_{\mathcal{J}}(u) \geq \ell$ and $d_{\mathcal{J}}(r) \geq d_{\mathcal{J}}(v_2) \geq \ell$. Since $d_{\mathcal{J}}(r) \leq \ell$ by hypothesis on r, we get $d_{\mathcal{J}}(r) = d_{\mathcal{J}}(u) = d_{\mathcal{J}}(v_2) = \ell$. Together with $r \leqslant_{\mathcal{J}} \delta(u)$ and $r \leqslant_{\mathcal{J}} \delta(v_2)$, this yields $r \mathrel{\mathcal{J}} \delta(u) \mathrel{\mathcal{J}} \delta(v_2)$, *i.e.*, $r \mathrel{\mathcal{J}} g \mathrel{\mathcal{J}} h$. This completes the proof that $g \mathrel{\mathcal{J}} h$. Since we also know that $hg = gh = h$, we have $h \leqslant_{\mathcal{R}} g$ and Lemma 1 yields $g \mathrel{\mathcal{R}} h$. We get $z \in N$ such that $g = hz$. Thus, we have $h = hg = hhz = hz = g$.

Altogether, we obtain $\delta(u) = \delta(v_1) = \delta(v_2) = g \in E(N)$. This implies that $(\beta(u), \beta(v_1))$ and $(\beta(u), \beta(v_2))$ are δ-pairs for β. Moreover, recall that $u = (u')^p$ where p is a multiple of $\omega(M)$. Hence, we have $\beta(u) \in E(M)$. Consequently, it follows from Lemma 23 that $\beta((uv_1uv_2u)^p) = \beta((uv_1uv_2u)^p uv_2u(uv_1uv_2u)^p)$. It now suffices to multiply by $\beta(x)$ on the left and $\beta(y)$ on the right to obtain $\beta(z_1) = \beta(z_2)$, as desired. ◀

We now turn to the second step of the proof, which is formalized in the following statement.

▶ **Proposition 26.** *Let $\ell \leq |N|$ and $s \in M$. There exists a $\mathrm{TL}[\mathcal{C}_\delta]$ formula $\varphi_{\ell,s}$ such that for every $w \in K_\ell$, we have $w, 0 \models \varphi_{\ell,s} \Leftrightarrow \beta(w) = s$.*

Let us first use Proposition 26 to complete the main proof: we have to show that every language recognized by β belongs to $\mathrm{TL}(\mathcal{C}_\delta)$. Clearly, it suffices to show that $\beta^{-1}(s) \in \mathrm{TL}(\mathcal{C}_\delta)$ for each $s \in M$. We apply Proposition 26 for $\ell = 0$. Since $K_0 = B^*$, this yields a formula $\varphi_{0,s} \in \mathrm{TL}[\mathcal{C}_\delta]$ such that $L_{min}(\varphi_{0,s}) = \beta^{-1}(s)$. Thus, $\beta^{-1}(s) \in \mathrm{TL}(\mathcal{C}_\delta)$, as desired.

It remains to prove Proposition 26. We construct $\varphi_{\ell,s} \in \mathrm{TL}[\mathcal{C}_\delta]$ by induction on $|N| - \ell$. If $\ell = |N|$, we define $\varphi_{\ell,s}$ so that $L_{min}(\varphi_{\ell,s}) = K_\ell \cap \beta^{-1}(s)$. Since $K_{|N|} \cap \beta^{-1}(s)$ is finite and $\mathrm{TL}[\mathcal{C}_\delta]$ is closed under disjunction, it suffices to build for every word $w \in B^*$ a $\mathrm{TL}[\mathcal{C}_\delta]$ formula φ_w defining $\{w\}$. Since $B^* \in \mathcal{C}_\delta$, one may use the "F" modality. For $w = b_1 \cdots b_n$, let

$\psi_w = \mathrm{F}\,(b_1 \wedge \mathrm{F}\,(b_2 \wedge \mathrm{F}\,(b_3 \wedge \cdots \wedge \mathrm{F}\,b_n)))$.

One may then choose $\varphi_w = \psi_w \wedge \bigwedge_{u \in B^*, |u|=|w|+1} \neg\psi_u$.

Assume now that $\ell < |N|$. We present a construction for splitting the words in K_ℓ into two parts: a prefix mapped to an element $r \in N$ such that $d_{\mathcal{J}}(r) \leq \ell$ (we handle it with Proposition 25) and a suffix that we abstract as a word in $K_{\ell+1}$ (we handle it by induction).

Let $w \in K_\ell$. For each position $i \in \mathsf{Pos}(w) \setminus \{0\}$ and $k \in \mathbb{N}$, we write $\sigma_k(w,i) \in B^*$ for the infix $w(i-1,j)$ where $j = \min(i+k, |w|+1)$. In other words, $\sigma_k(w,i) = w[i] \cdots w[i+k-1]$ if $i+k-1 \leq |w|$ and $\sigma_k(w,i) = w[i] \cdots w[|w|]$ otherwise. In particular, we have $|\sigma_k(w,i)| \leq k$.

▶ **Lemma 27.** *Let $k \leq \ell + 1$ and $u \in B^*$ be such that $|u| \leq k$. There exists a formula $\pi_{k,u} \in \mathrm{TL}[\mathcal{C}_\delta]$ such that for all $w \in K_\ell$ and $i \in \mathsf{Pos}(w) \setminus \{0\}$, $w, i \models \pi_{k,u} \Leftrightarrow \sigma_k(w,i) = u$.*

Proof. If $u = \varepsilon$, it suffices to define $\pi_{k,u} = \top$ when $k = 0$ and $\pi_{k,u} = max$ when $k \geq 1$. Assume now that $|u| \geq 1$. If $k \leq 1$, it follows that $|u| = 1 = k$. Hence, u is a letter $b \in B$ and it suffices to define $\pi_{k,u} = b$. Assume now that $k \geq 2$. Let $C \subseteq B$ be the set of letters mapped to 1_N under δ, so that $H \stackrel{\text{def}}{=} \delta^{-1}(1_N) = C^*$. By definition, $H \in \mathcal{C}_\delta$. Since $2 \leq k \leq \ell + 1$, we have $\ell \geq 1$, which implies, by definition of K_ℓ, that no word of K_ℓ can contain a letter b with $d_{\mathcal{J}}(b) = 0$. In particular, words of K_ℓ cannot contain letters of C. Therefore, if $w \in K_\ell$, $i \in \mathsf{Pos}(w)$ and $\psi \in \mathrm{TL}[\mathcal{C}_\delta]$, we have $w, i \models \mathrm{F}_H\,\psi$ if and only if $w, i+1 \models \psi$. Let $u = b_1 \cdots b_n$ (with $b_i \in B$). We have $n = |u| \leq k$ by hypothesis. We consider two cases for defining $\pi_{k,u}$:

- If $n = k$, we let $\pi_{k,u} = (b_1 \wedge \mathrm{F}_H\,(b_2 \wedge \mathrm{F}_H\,(b_3 \wedge \cdots \mathrm{F}_H\,b_n)))$.
- If $n < k$, we let $\pi_{k,u} = (b_1 \wedge \mathrm{F}_H\,(b_2 \wedge \mathrm{F}_H\,(b_3 \wedge \cdots \mathrm{F}_H\,(b_n \wedge \mathrm{F}_H\,max))))$.

The above fact on F_H implies that this definition fulfills the desired property. ◀

Pointed positions. Consider $w \in K_\ell$. We say that an arbitrary position $i \in \mathsf{Pos}(w)$ is *pointed* when either $i \in \{0, |w|+1\}$, or $i \in \mathsf{Pos}_c(w)$ and $d_{\mathcal{J}}(\sigma_{\ell+1}(w,i)) \geq \ell + 1$.

▶ **Definition 28** (Detection of pointed positions in $\mathrm{TL}[\mathcal{C}_\delta]$)**.** *Let $\pi = min \vee max \vee \bigvee_{u \in U} \pi_{\ell+1,u}$ where $U = \{u \in B^* \mid |u| \leq \ell + 1 \text{ and } d_{\mathcal{J}}(u) \geq \ell + 1\}$. By definition of $\pi_{\ell+1,u}$ in Lemma 27, we know that for $w \in K_\ell$ and $i \in \mathsf{Pos}(w)$, we have $w, i \models \pi$ if and only if position i is pointed.*

A position $i \in \mathsf{Pos}(w)$ which is *not* pointed is said to be *safe*. We now prove that we may constrain the evaluation of $\mathrm{TL}[\mathcal{C}_\delta]$ formulas to infixes that only contain *safe* positions.

▶ **Lemma 29.** *Let $\psi \in \mathrm{TL}[\mathcal{C}_\delta]$ and $H \in \mathcal{C}_\delta$. There exist two formulas $\mathrm{F}^{sa}_H\,\psi$ and $\mathrm{P}^{sa}_H\,\psi$ of $\mathrm{TL}[\mathcal{C}_\delta]$ such that for all $w \in K_\ell$ and all $i \in \mathsf{Pos}(w)$, the two following properties hold:*

- *$w, i \models \mathrm{F}^{sa}_H\,\psi$ if and only if there exists $j \in \mathsf{Pos}(w)$ such that $j > i$, $w, j \models \psi$, $w(i,j) \in H$ and all positions $h \in \mathsf{Pos}(w)$ such that $i < h < j$ are safe.*
- *$w, i \models \mathrm{P}^{sa}_H\,\psi$ if and only if there exists $j \in \mathsf{Pos}(w)$ such that $j < i$, $w, j \models \psi$, $w(j,i) \in H$ and all positions $h \in \mathsf{Pos}(w)$ such that $j < h < i$ are safe.*

Proof. We begin by characterizing infixes containing only safe positions. Let $w \in K_\ell$ and $i, j \in \mathsf{Pos}(w)$ be such that $i < j$. We prove that the following two properties are equivalent:

1. All positions $h \in \mathsf{Pos}(w)$ such that $i < h < j$ are safe.
2. Either $\delta(w(i,j)) = 1_N$ or $d_{\mathcal{J}}(w(i,j)\sigma_\ell(w,j)) \leq \ell$.

Assume first that all positions $h \in \mathsf{Pos}(w)$ such that $i < h < j$ are safe. If $i + 1 = j$, then $w(i,j) = \varepsilon$, whence $\delta(w(i,j)) = 1_N$. Assume now that $i+1 < j$. Observe that $w(i,j)\sigma_\ell(w,j)$ belongs to K_ℓ since it is an infix of $w \in K_\ell$. Moreover, since $i+1 < j$, there exists at least one $h \in \mathsf{Pos}(w)$ such that $i < h < j$. Combined with the assumption that all such positions h are safe, this implies that for every $x, y, z \in B^*$ such that $xyz = w(i,j)\sigma_\ell(w,j)$ and $|y| \leq \ell + 1$, we have $d_{\mathcal{J}}(y) \leq \ell$. Therefore, Lemma 24 entails that $d_{\mathcal{J}}(w(i,j)\sigma_\ell(w,j)) \leq \ell$, as desired.

Conversely, assume that either $\delta(w(i,j)) = 1_N$ or $d_{\mathcal{J}}(w(i,j)\sigma_\ell(w,j)) \leq \ell$. We start with the latter case. Since $w(i,j)\sigma_\ell(w,j) \in K_\ell$, Lemma 24 implies that for every $x,y,z \in B^*$ such that $xyz = w(i,j)\sigma_\ell(w,j)$ and $|y| \leq \ell+1$, we have $d_{\mathcal{J}}(y) \leq \ell$. In particular, it follows that every $h \in \mathsf{Pos}(w)$ such that $i < h < j$ is safe. Assume now that $\delta(w(i,j)) = 1_N$. If $\ell = 0$, then $\sigma_\ell(w,j) = \varepsilon$ and we are back to the previous case. Otherwise, $\ell \geq 1$ and since $w \in K_\ell$, the fact that $\delta(w(i,j)) = 1_N$ yields $w(i,j) = \varepsilon$, which completes the argument.

We are now ready to complete the proof of the lemma. Let $\psi \in \mathrm{TL}[\mathcal{C}_\delta]$ and $H \in \mathcal{C}_\delta$. For every $r \in N$, we let $H_r = H \cap \delta^{-1}(r)$ and $U_r = \{u \in B^* \mid |u| \leq \ell \text{ and } d_{\mathcal{J}}(r\delta(u)) \leq \ell\}$. Observe that $H_r \in \mathcal{C}_\delta$. Now, in view of the preliminary result, it suffices to define,

$$\mathrm{F}^{sa}_H\,\psi = \mathrm{F}_{H_{1_N}}\,\psi \vee \bigvee_{r\in N}\bigvee_{u\in U_r} \mathrm{F}_{H_r}\,(\pi_{\ell,u} \wedge \psi) \quad\text{and}\quad \mathrm{P}^{sa}_H\,\psi = \mathrm{P}_{H_{1_N}}\,\psi \vee \bigvee_{r\in N}\bigvee_{u\in U_r} (\pi_{\ell,u} \wedge \mathrm{P}_{H_r}\,\psi)\,.$$

This completes the proof. ◀

Pointed decomposition. Let $w \in K_\ell$ and let $0 = i_0 < i_1 < \cdots < i_n < i_{n+1} = |w|+1$ be all the pointed positions of w. The *pointed decomposition* of w is the decomposition $w = w_0b_1w_1\cdots b_nw_n$ where the highlighted letters $b_1,\dots,b_n \in B$ are those carried by the pointed positions $i_1,\dots,i_n$. For $0 \leq j \leq n$, we associate the word $f(w,i_j) = w_j$ to the pointed position i_j. Moreover, we define a new word $\widehat{w} \in B^*$ built from the suffix $b_1w_1\cdots b_nw_n$. For $1 \leq j \leq n$, let $(t_j,q_j) = (\beta(b_jw_j),\delta(b_jw_j)) \in P$. By definition of β and δ, we know that there is a letter $b_{t_j,q_j} \in B$ such that $(\beta(b_{t_j,q_j}),\delta(b_{t_j,q_j})) = (t_j,q_j)$. We let $\widehat{w} = b_{t_1,q_1}\cdots b_{t_n,q_n}$. Note that by definition, $\beta(b_1w_1\cdots b_nw_n) = \beta(\widehat{w})$ and $\delta(b_1w_1\cdots b_nw_n) = \delta(\widehat{w})$. Finally, we define a surjective map $i \mapsto \mu(i)$ associating a position $\mu(i) \in \mathsf{Pos}(\widehat{w})$ to each *pointed* position $i \in \mathsf{Pos}(w)$: for $0 \leq j \leq n+1$, we let $\mu(i_j) = j$. We complete this definition with a key property. For every pointed position $i \in \{0\} \cup \mathsf{Pos}_c(w)$, one can compute the images of the word $f(w,i)$ under β and δ with a $\mathrm{TL}[\mathcal{C}_\delta]$ formula. This is where we use Proposition 25.

▶ **Lemma 30.** *Let $(t,r) \in M \times N$. There exists $\Gamma_{t,r} \in \mathrm{TL}[\mathcal{C}_\delta]$ such that for all $w \in K_\ell$ and all pointed positions $i \in \{0\} \cup \mathsf{Pos}_c(w)$, we have $w,i \models \Gamma_{t,r} \Leftrightarrow \beta(f(w,i)) = t$ and $\delta(f(w,i)) = r$.*

Proof. First observe that by definition, if $w \in K_\ell$ and $i \in \{0\} \cup \mathsf{Pos}_c(w)$ is pointed, the infix $f(w,i)$ contains only *safe* positions. Hence, for every $x,y,z \in B^*$ such that $f(w,i) = xyz$ and $|y| \leq \ell+1$, we have $d_{\mathcal{J}}(y) \leq \ell$. By Lemma 24, it follows that $d_{\mathcal{J}}(f(w,i)) \leq \ell$. Therefore, if $d_{\mathcal{J}}(r) > \ell$, then $\delta(f(w,i))$ cannot be equal to r, and it suffices to define $\Gamma_{t,r} = \bot$.

We now assume that $d_{\mathcal{J}}(r) \leq \ell$. Proposition 25 implies that $K_\ell \cap \beta^{-1}(t) \cap \delta^{-1}(r) \in \mathrm{TL}(\mathcal{C}_\delta)$. We get a formula $\psi \in \mathrm{TL}[\mathcal{C}_\delta]$ such that for every $u \in K_\ell$, we have $u,0 \models \psi$ if and only if $\beta(u) = t$ and $\delta(u) = r$. Using Lemma 29, we modify ψ so that given $w \in K_\ell$, the evaluation of ψ at a pointed position i is constrained to the infix $f(w,i)$. More precisely, we use structural induction to build two formulas $\langle\psi\rangle_{min}$ and $\langle\psi\rangle_{max}$ such that given $w \in K_\ell$, a pointed position $i \in \{0\} \cup \mathsf{Pos}_c(w)$ and $j \in \mathsf{Pos}(f(w,i))$, the two following properties hold:

- If $j \leq |f(w,i)|$, then $w,i+j \models \langle\psi\rangle_{min} \Leftrightarrow f(w,i),j \models \psi$.
- If $1 \leq j$, then $w,i+j \models \langle\psi\rangle_{max} \Leftrightarrow f(w,i),j \models \psi$.

It will then suffice to define $\Gamma_{t,r} = \langle\psi\rangle_{min}$. We only describe the construction, and leave it to the reader to check that it satisfies the above properties. Note that we use the formula $\pi \in \mathrm{TL}[\mathcal{C}_\delta]$ of Definition 28 that detects pointed positions.

For $\psi \in B \cup \{\top,\bot\}$, we let $\langle\psi\rangle_{min} = \langle\psi\rangle_{max} = \psi$. If $\psi = min$, we let $\langle\psi\rangle_{min} = \pi$ and $\langle\psi\rangle_{max} = \bot$. If $\psi = max$, we let $\langle\psi\rangle_{min} = \bot$ and $\langle\psi\rangle_{max} = \pi$. We handle Boolean operators in the expected way. For instance, we define $\langle\psi' \vee \psi''\rangle_{min} = \langle\psi'\rangle_{min} \vee \langle\psi''\rangle_{min}$, $\langle\psi' \wedge \psi''\rangle_{min} = \langle\psi'\rangle_{min} \wedge \langle\psi''\rangle_{min}$ and $\langle\neg\psi'\rangle_{min} = \neg\langle\psi'\rangle_{min}$, and similarly for $\langle\cdot\rangle_{max}$.

If $\psi = \mathrm{F}_H\,\psi'$ for $H \in \mathcal{C}_\delta$, we let $\langle\psi\rangle_{min} = \mathrm{F}^{sa}_H\,\langle\psi'\rangle_{max}$ and $\langle\psi\rangle_{max} = \neg\pi \wedge \mathrm{F}^{sa}_H\,\langle\psi'\rangle_{max}$. Symmetrically, if $\psi = \mathrm{P}_H\,\psi'$ for some $H \in \mathcal{C}_\delta$, we define $\langle\psi\rangle_{min} = \neg\pi \wedge \mathrm{P}^{sa}_H\,\langle\psi'\rangle_{min}$ and $\langle\psi\rangle_{max} = \mathrm{P}^{sa}_H\,\langle\psi'\rangle_{min}$. This concludes the inductive construction of $\langle\psi\rangle_{min}$ and $\langle\psi\rangle_{max}$ and the proof of the proposition. ◀

Construction of the formulas $\varphi_{\ell,s}$. We are ready to complete the proof of Proposition 26. For every $s \in M$, we build a formula $\zeta_s \in \mathrm{TL}[\mathcal{C}_\delta]$ such that for every $w \in K_\ell$, we have $w, 0 \models \zeta_s \Leftrightarrow \beta(\widehat{w}) = s$. Given $s \in M$, it will then suffice to define $\varphi_{\ell,s} \in \mathrm{TL}[\mathcal{C}_\delta]$ as follows:

$$\varphi_{\ell,s} = \bigvee_{\{(s_1,s_2)\in M^2 \mid s_1 s_2 = s\}} \Big(\Big(\bigvee_{r\in N} \Gamma_{s_1,r}\Big) \wedge \zeta_{s_2}\Big).$$

Indeed, it is straightforward that for every word $w \in K_\ell$, we have $\beta(w) = \beta(f(w,0))\beta(\widehat{w})$. Consequently, by definition of $\varphi_{\ell,s}$, we get $w, 0 \models \varphi_{\ell,s} \Leftrightarrow \beta(f(w,0))\beta(\widehat{w}) = s \Leftrightarrow \beta(w) = s$ for all $w \in K_\ell$, which concludes the proof of Proposition 26. We now concentrate on building ζ_s. This is where we use induction in Proposition 26. Indeed, we have the following lemma.

▶ **Lemma 31.** *For every $w \in K_\ell$, we have $\widehat{w} \in K_{\ell+1}$.*

Proof. Let $k \leq \ell + 1$ and $x, y, z \in B^*$ such that $\widehat{w} = xyz$ and $|y| = k$. We have to prove that $d_\beth(y) \geq k$. Let $w = w_0 b_1 w_1 \cdots b_n w_n$ be the pointed decomposition of w. By definition of $\widehat{w}$, we have $\delta(y) = \delta(b_h w_h \cdots b_{h+k-1} w_{h+k-1})$ for some $h \leq n$. Let $u = b_h w_h \cdots b_{h+k-1} w_{h+k-1}$. We have to show that $d_\beth(y) = d_\beth(u) \geq k$. Clearly, $|u| \geq k$. Hence, if $k \leq \ell$, the hypothesis that $w \in K_\ell$ yields $d_\beth(u) \geq k$. Otherwise, $k = \ell + 1$. Thus, $|u| \geq \ell + 1$ and since the position labeled by b_h in w is pointed, this yields $d_\beth(u) \geq \ell + 1$. In both cases, we get $d_\beth(y) \geq k$. ◀

Let $s \in M$. In view of Lemma 31, induction on $|N| - \ell$ in Proposition 26 yields a $\mathrm{TL}[\mathcal{C}_\delta]$ formula ψ_s such that for every $w \in K_\ell$, we have $\widehat{w}, 0 \models \psi_s \Leftrightarrow \beta(\widehat{w}) = s$. Thus, it now suffices to prove that for every $\psi \in \mathrm{TL}[\mathcal{C}_\delta]$, there exists a formula $\lfloor\psi\rfloor \in \mathrm{TL}[\mathcal{C}_\delta]$ such that for every $w \in K_\ell$ and every pointed position $i \in \mathsf{Pos}(w)$, we have $w, i \models \lfloor\psi\rfloor \Leftrightarrow \widehat{w}, \mu(i) \models \psi$. It will then follow, for $i = 0$, that $w, 0 \models \lfloor\psi_s\rfloor \Leftrightarrow \beta(\widehat{w}) = s$, meaning that we can define $\zeta_s = \lfloor\psi_s\rfloor$.

We construct $\lfloor\psi\rfloor$ by structural induction on ψ. If $\psi \in \{min, max, \top, \bot\}$, we let $\lfloor\psi\rfloor = \psi$. Suppose now that $\psi = b_{t,q} \in B$ for $(t,q) \in P$. Thus, when evaluated in w at a pointed position i carrying a "b", we want $\lfloor\psi\rfloor$ to check that $\beta(b)\beta(f(w,i)) = t$ and $\delta(b)\delta(f(w,i)) = q$. Let $T = \big\{(b,t',q') \in B \times M \times N \mid \beta(b)t' = t \text{ and } \delta(b)q' = q\big\}$. Using the formulas $\Gamma_{t',q'}$ from Lemma 30, we define $\psi = \bigvee_{(b,t',q')\in T} (b \wedge \Gamma_{t',q'})$. Boolean operators are handled as expected. It remains to deal with temporal modalities, *i.e.*, the case where there exists $H \in \mathcal{C}_\delta$ such that $\psi = \mathrm{F}_H\,\psi'$ or $\psi = \mathrm{P}_H\,\psi'$. For every $b \in B$, let $F_b = \big\{r \in N \mid \delta(b)r \in \delta(H)\big\}$. We define:

$$\lfloor \mathrm{F}_H\,\psi' \rfloor \stackrel{\text{def}}{=} \begin{cases} \mathrm{F}^{sa}_{B^*}\,\big(\pi \wedge \big(\bigvee_{b\in B}\big(b \wedge \mathrm{F}_{\delta^{-1}(F_b)}\,(\pi \wedge \lfloor\psi'\rfloor)\big)\big)\big) & \text{if } \varepsilon \notin H,\\ \mathrm{F}^{sa}_{B^*}\,\big(\pi \wedge \big(\bigvee_{b\in B}\big(b \wedge \mathrm{F}_{\delta^{-1}(F_b)}\,(\pi \wedge \lfloor\psi'\rfloor)\big) \vee \lfloor\psi'\rfloor\big)\big) & \text{if } \varepsilon \in H.\end{cases}$$

$$\lfloor \mathrm{P}_H\,\psi' \rfloor \stackrel{\text{def}}{=} \begin{cases} \bigvee_{b\in B} \mathrm{P}_{\delta^{-1}(F_b)}\,(\pi \wedge b \wedge \mathrm{P}^{sa}_{B^*}\,(\pi \wedge \lfloor\psi'\rfloor)) & \text{if } \varepsilon \notin H,\\ \mathrm{P}^{sa}_{B^*}\,(\pi \wedge \lfloor\psi'\rfloor) \vee \bigvee_{b\in B} \mathrm{P}_{\delta^{-1}(F_b)}\,(\pi \wedge b \wedge \mathrm{P}^{sa}_{B^*}\,(\pi \wedge \lfloor\psi'\rfloor)) & \text{if } \varepsilon \in H.\end{cases}$$

We give an intuition when $\psi = \mathrm{F}_H\,\psi'$ and $\varepsilon \notin H$. Let $w_0 b_1 w_1 \cdots b_n w_n$ be the pointed decomposition of w and $\widehat{w} = b'_1 \cdots b'_n$. Let $i_k \in \mathsf{Pos}(w)$ be the position of the distinguished b_k, so that $\mu(i_k) = k$. Now, $\widehat{w}, k \models \mathrm{F}_H\,\psi'$ when there exists $m > k$ such that $\widehat{w}, m \models \psi'$ and $b'_{k+1} \cdots b'_{m-1} \in H$. The construction ensures that $w, i_k \models \lfloor \mathrm{F}_H\,\psi' \rfloor$ when there exists $m > k$ such that $w, i_m \models \lfloor\psi'\rfloor$ and $b_{k+1} w_{k+1} \cdots b_{m-1} w_{m-1} \in H$. The purpose of using $\mathrm{F}^{sa}_{B^*}\,(\pi \wedge \dots)$ is to "jump" to b_{k+1}. The remainder checks that the next jump, to a pointed position, determines a word of $\delta^{-1}(\delta(H)) = H$. More generally, one can check that $w, i \models \lfloor\psi\rfloor \Leftrightarrow \widehat{w}, \mu(i) \models \psi$ for all $w \in K_\ell$ and all pointed positions $i \in \mathsf{Pos}(w)$. This concludes the proof.

6 Natural restrictions of generalized unary temporal logic

We turn to two natural restrictions of the classes $\mathrm{TL}(\mathcal{C})$, which were defined in [26]: the pure-future and pure-past fragments. For a class $\mathcal{C}$, we write $\mathrm{FL}[\mathcal{C}] \subseteq \mathrm{TL}[\mathcal{C}]$ for the set of all formulas that contain only *future* modalities (*i.e.*, the modalities P_L are disallowed). Symmetrically, $\mathrm{PL}[\mathcal{C}] \subseteq \mathrm{TL}[\mathcal{C}]$ is the set of all formulas in $\mathrm{TL}[\mathcal{C}]$ that contain only *past* modalities (*i.e.*, the modalities F_L are disallowed).

We now define the two associated operators $\mathcal{C} \mapsto \mathrm{FL}(\mathcal{C})$ and $\mathcal{C} \mapsto \mathrm{PL}(\mathcal{C})$. For every class $\mathcal{C}$, let $\mathrm{FL}(\mathcal{C})$ be the class consisting of all languages $L_{min}(\varphi)$ where $\varphi \in \mathrm{FL}[\mathcal{C}]$. Symmetrically, we write $\mathrm{PL}(\mathcal{C})$ for the class consisting of all languages $L_{max}(\varphi)$, with $\varphi \in \mathrm{PL}[\mathcal{C}]$.

▶ Remark 32. Note that $\mathrm{FL}[\mathcal{C}]$ formulas are evaluated at the leftmost unlabeled position whereas $\mathrm{PL}[\mathcal{C}]$ formulas are evaluated at the rightmost unlabeled position.

6.1 Connection with left and right polynomial closure

The main ideas to establish decidable characterizations for $\mathrm{FL}(\mathcal{C})$ and $\mathrm{PL}(\mathcal{C})$ follow the lines of the proof of Theorem 12. However, there are some differences. First, for the easy direction (proving that some property on $\mathcal{C}$-orbits is necessary), we have to adapt Lemma 20 to the operators $\mathcal{C} \mapsto \mathrm{FL}(\mathcal{C})$ and $\mathcal{C} \mapsto \mathrm{PL}(\mathcal{C})$. We prove these adapted properties in the extended version of this paper [29] as corollaries of results presented in [26].

The proof of the difficult direction is mostly identical to that in Theorem 12. However, there is a key difference: we have to find a substitute for Proposition 25, whose proof relied the inclusion $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C})) \subseteq \mathrm{TL}(\mathcal{C})$ from Proposition 10. We replace unambiguous polynomial closure (UPol) by two variants, called *right* and *left* polynomial closure (RPol and LPol). It is shown [26] that $\mathrm{RPol}(\mathrm{BPol}(\mathcal{C})) \subseteq \mathrm{FL}(\mathcal{C})$ and $\mathrm{LPol}(\mathrm{BPol}(\mathcal{C})) \subseteq \mathrm{PL}(\mathcal{C})$ for every prevariety $\mathcal{C}$: this serves as a substitute for Proposition 10. Finally, while no simple generic characterization of the classes $\mathrm{RPol}(\mathrm{BPol}(\mathcal{C}))$ and $\mathrm{LPol}(\mathrm{BPol}(\mathcal{C}))$ are known, we are able to replace Theorem 11 by combining independent characterizations of the operators Pol and RPol (resp. Pol and LPol) from [23, 21].

We now establish a connection between the operators $\mathcal{C} \mapsto \mathrm{FL}(\mathcal{C})$ and $\mathcal{C} \mapsto \mathrm{PL}(\mathcal{C})$ and the two weaker variants RPol and LPol of unambiguous polynomial closure. Consider a marked product $L_0a_1L_1 \cdots a_nL_n$. For $1 \leq i \leq n$, we write $H_i = L_1a_1L_2 \cdots a_{i-1}L_{i-1}$ and $K_i = L_ia_{i+1}L_{i+1} \cdots a_nL_n$. We say that $L_0a_1L_1 \cdots a_nL_n$ is *right deterministic* (resp. *left deterministic*) when we have $A^*a_iK_i \cap K_i = \emptyset$ (resp. $H_ia_iA^* \cap H_i = \emptyset$) for every $i \leq n$. The *right polynomial closure* of a class $\mathcal{C}$, written $\mathrm{RPol}(\mathcal{C})$, consists of all *finite disjoint unions* of *right deterministic marked products* $L_0a_1L_1 \cdots a_nL_n$ such that $L_0, \dots, L_n \in \mathcal{C}$ (by "disjoint" we mean that the languages in the union must be pairwise disjoint). Similarly, the *left polynomial closure* $\mathrm{LPol}(\mathcal{C})$ of $\mathcal{C}$ consists of all finite disjoint unions of *left* deterministic marked products $L_0a_1L_1 \cdots a_nL_n$ such that $L_0, \dots, L_n \in \mathcal{C}$. While this is not immediate, it is known [21] that when the input class $\mathcal{C}$ is a prevariety, then so are $\mathrm{RPol}(\mathcal{C})$ and $\mathrm{LPol}(\mathcal{C})$.

As expected, we are interested in the "combined" operators $\mathcal{C} \mapsto \mathrm{RPol}(\mathrm{BPol}(\mathcal{C}))$ and $\mathcal{C} \mapsto \mathrm{LPol}(\mathrm{BPol}(\mathcal{C}))$. Indeed, the first one is connected to the classes $\mathrm{FL}(\mathcal{C})$ by the following result proved in [26, Proposition 5].

▶ **Proposition 33.** *For every prevariety $\mathcal{C}$, we have* $\mathrm{RPol}(\mathrm{BPol}(\mathcal{C})) \subseteq \mathrm{FL}(\mathcal{C})$.

We have the following symmetrical statement for $\mathrm{PL}(\mathcal{C})$.

▶ **Proposition 34.** *For every prevariety $\mathcal{C}$, we have* $\mathrm{LPol}(\mathrm{BPol}(\mathcal{C})) \subseteq \mathrm{PL}(\mathcal{C})$.

Propositions 33 and 34 serve as the replacement of Proposition 10 when dealing with the classes $\mathrm{FL}(\mathcal{C})$ and $\mathrm{PL}(\mathcal{C})$, respectively. It now remains to replace the generic algebraic characterization of the classes $\mathrm{UPol}(\mathrm{BPol}(\mathcal{C}))$ presented in Theorem 11. This is more tricky as no such characterization is known for the classes $\mathrm{RPol}(\mathrm{BPol}(\mathcal{C}))$ (nor for the classes $\mathrm{LPol}(\mathrm{BPol}(\mathcal{C}))$). Yet, we manage to prove a *sufficient condition* for a language to belong to $\mathrm{RPol}(\mathrm{BPol}(\mathcal{C}))$ or $\mathrm{LPol}(\mathrm{BPol}(\mathcal{C}))$ by combining results of [27] and [21]. While it does *not* characterize these classes in general, it suffices for our needs: proving that particular languages belong to $\mathrm{RPol}(\mathrm{BPol}(\mathcal{C}))$ (and therefore to $\mathrm{FL}(\mathcal{C})$ by Proposition 33) or to $\mathrm{LPol}(\mathrm{BPol}(\mathcal{C}))$ (and therefore to $\mathrm{FL}(\mathcal{C})$ by Proposition 34).

▶ **Proposition 35.** *Let $\mathcal{C}$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \to M$ be its syntactic morphism. Assume that α satisfies the following property:*

$$(esete)^{\omega+1} = ete(esete)^{\omega} \quad \text{for every } \mathcal{C}\text{-pair } (e, s) \in M^2 \text{ and every } t \in M. \tag{5}$$

Then, $L \in \mathrm{RPol}(\mathrm{BPol}(\mathcal{C}))$.

▶ **Proposition 36.** *Let $\mathcal{C}$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \to M$ be its syntactic morphism. Assume that α satisfies the following property:*

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

Then, $L \in \mathrm{LPol}(\mathrm{BPol}(\mathcal{C}))$.

Since Propositions 35 and 36 are symmetrical, we only prove the first one and leave the second to the reader.

Proof of Proposition 35. We use a generic characterization of the classes $\mathrm{RPol}(\mathcal{D})$ proved in [21]. Let us first present it. For every class $\mathcal{D}$, we define a preorder $\preceq_{\mathcal{D}}$ and an equivalence $\sim_{\mathcal{D}}$ over M. Given $s, t \in M$, we let,

$$\begin{array}{lll} s \sim_{\mathcal{D}} t & \text{if and only if} & s \in F \Leftrightarrow t \in F \text{ for every } F \subseteq M \text{ such that } \alpha^{-1}(F) \in \mathcal{D}, \\ s \preceq_{\mathcal{D}} t & \text{if and only if} & s \in F \Rightarrow t \in F \text{ for every } F \subseteq M \text{ such that } \alpha^{-1}(F) \in \mathcal{D}. \end{array}$$

Clearly, $\preceq_{\mathcal{D}}$ is a preorder on M and $\sim_{\mathcal{D}}$ is the equivalence generated by $\preceq_{\mathcal{D}}$. When $\alpha : A^* \to M$ is the syntactic morphism of L, it is shown in [21, Theorem 4.1] that for every prevariety $\mathcal{D}$, we have $L \in \mathrm{RPol}(\mathcal{D})$ if and only if $s^{\omega+1} = ts^{\omega}$ for all $s, t \in M$ such that $s \sim_{\mathcal{D}} t$.

Hence, since $\mathrm{BPol}(\mathcal{C})$ is a prevariety, it suffices to prove that for every $s, t \in M$ such that $s \sim_{\mathrm{BPol}(\mathcal{C})} t$, we have $s^{\omega+1} = ts^{\omega}$. We fix s, t for the proof. Since $s \sim_{\mathrm{BPol}(\mathcal{C})} t$, we have $s \preceq_{\mathrm{BPol}(\mathcal{C})} t$. Moreover, let *co*-$\mathrm{Pol}(\mathcal{C})$ be the class consisting of all complements of languages in $\mathrm{Pol}(\mathcal{C})$ (*i.e.*, $L \in$ *co*-$\mathrm{Pol}(\mathcal{C})$ if and only if $A^* \setminus L \in$ *co*-$\mathrm{Pol}(\mathcal{C})$). Clearly, we have *co*-$\mathrm{Pol}(\mathcal{C}) \subseteq \mathrm{BPol}(\mathcal{C})$. Hence, the definition implies that $s \preceq_{co\text{-}\mathrm{Pol}(\mathcal{C})} t$

Moreover, it is shown in [27, Lemma 6.6] that $\preceq_{co\text{-}\mathrm{Pol}(\mathcal{C})}$ is the least preorder on M such that for every $x, y, q \in M$ and $e \in E(M)$, if $(e, q) \in M^2$ is a $\mathcal{C}$-pair, then $xeqey \preceq_{co\text{-}\mathrm{Pol}(\mathcal{C})} xey$ (the proof is based on the algebraic characterization of $\mathrm{Pol}(\mathcal{C})$, see [23]). This yields $s_0, \dots, s_n \in M$ such that $s = s_0$, $t = s_n$ and, for every $i \leq n$, there exist $x, y, q \in M$ and $e \in E(M)$ such that $(e, q) \in M^2$ is a $\mathcal{C}$-pair, $s_{i-1} = xeqey$ and $s_i = xey$. We use induction on i to prove that $s^{\omega+1} = s_i s^{\omega}$ for every $i \leq n$. Since $s_n = t$, the case $i = n$ yields the desired result. When $i = 0$, it is immediate that $s^{\omega+1} = s_0 s^{\omega}$ since $s_0 = s$. Assume now that $i \geq 1$.

By induction hypothesis, we know that $s^{\omega+1} = s_{i-1}s^{\omega}$. Moreover, we have $x, y, q \in M$ and $e \in E(M)$ such that $(e, q) \in M^2$ is a $\mathcal{C}$-pair, $s_{i-1} = xeqey$ and $s_i = xey$. Since $(s^{\omega+1})^{\omega+2} = s^{\omega+2}$, we get $s^{\omega+2} = (xeqeys^{\omega})^{\omega+2}$. Hence, we get

$$\begin{aligned} s^{\omega+2} &= x\ (eqeys^{\omega}xe)^{\omega+1}\ eqeys^{\omega} \\ &= x\ eys^{\omega}xe(eqeys^{\omega}xe)^{\omega}\ eqeys^{\omega} \quad \text{by (5) since } (e,q) \text{ is a } \mathcal{C}\text{-pair} \\ &= xeys^{\omega}(xeqeys^{\omega})^{\omega+1}. \end{aligned}$$

This yields, $s^{\omega+2} = s_i s^{\omega}(s_{i-1}s^{\omega})^{\omega+1} = s_i s^{\omega}(s^{\omega+1})^{\omega+1} = s_i s^{\omega+1}$. It now remains to multiply by $s^{\omega-1}$ on the right to get $s^{\omega+1} = s_i s^{\omega}$, as desired. ◀

6.2 Statements

The classes FL($\mathcal{C}$) and PL($\mathcal{C}$) admit algebraic characterizations similar to that of TL($\mathcal{C}$). We reuse the $\mathcal{C}$-orbits introduced in Section 3. Let $\mathcal{X} \in \{\mathcal{L}, \mathcal{R}, \mathcal{J}\}$ be one the Green relations defined in Section 2. A monoid M is *$\mathcal{X}$-trivial* when $s\ \mathcal{X}\ t$ implies $s = t$ for all $s, t \in M$. It is standard and simple to verify that a finite monoid M is $\mathcal{R}$-trivial (resp. $\mathcal{L}$-trivial) if and only if for all $s, t \in M$, we have $(st)^{\omega}s = (st)^{\omega}$ (resp. $t(st)^{\omega} = (st)^{\omega}$), see [17, 20] for a proof. We are now able to present the two symmetrical characterizations of FL($\mathcal{C}$) and PL($\mathcal{C}$).

▶ **Theorem 37.** *Let $\mathcal{C}$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \to M$ be its syntactic morphism. The two following properties are equivalent:*
1. *$L \in$ FL($\mathcal{C}$).*
2. *Every $\mathcal{C}$-orbit for α is $\mathcal{L}$-trivial.*

▶ **Theorem 38.** *Let $\mathcal{C}$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \to M$ be its syntactic morphism. The two following properties are equivalent:*
1. *$L \in$ PL($\mathcal{C}$).*
2. *Every $\mathcal{C}$-orbit for α is $\mathcal{R}$-trivial.*

Since FL($\mathcal{C}$) and PL($\mathcal{C}$) are symmetrical, it is natural to consider a third class denoted FL($\mathcal{C}$) ∩ PL($\mathcal{C}$). It consists of all languages belonging simultaneously to FL($\mathcal{C}$) and PL($\mathcal{C}$). It is standard that the finite monoids which are both $\mathcal{L}$-trivial and $\mathcal{R}$-trivial are exactly the $\mathcal{J}$-trivial monoids (see [17, 20]). This yields the following corollary of Theorems 37 and 38.

▶ **Corollary 39.** *Let $\mathcal{C}$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \to M$ be its syntactic morphism. The two following properties are equivalent:*
1. *$L \in$ FL($\mathcal{C}$) ∩ PL($\mathcal{C}$).*
2. *Every $\mathcal{C}$-orbit for α is $\mathcal{J}$-trivial.*

Recall that given a regular language $L \subseteq A^*$ as input, its syntactic morphism $\alpha : A^* \to M$ can be computed. Moreover, Lemma 5 implies that all $\mathcal{C}$-orbits for α can be computed when $\mathcal{C}$-separation is decidable. Thus, the three above characterizations yield the following corollary.

▶ **Corollary 40.** *Let $\mathcal{C}$ be a prevariety with decidable separation. Then, the classes* FL($\mathcal{C}$), PL($\mathcal{C}$) *and* FL($\mathcal{C}$) ∩ PL($\mathcal{C}$) *have decidable membership.*

We leave the proof of Theorem 37 for the extended version of this paper [29] (it omits the proof of Theorem 38, which is symmetrical).

7 Conclusion

We presented generic characterizations of the classes $\mathrm{TL}(\mathcal{C})$, $\mathrm{FL}(\mathcal{C})$ and $\mathrm{PL}(\mathcal{C})$. While the proofs are complex, the statements are simple and elegant. They generalize in a natural way all known characterizations of classes built with these operators. As a corollary, we obtained that if $\mathcal{C}$ is a prevariety with decidable *separation*, then all classes $\mathrm{TL}(\mathcal{C})$, $\mathrm{FL}(\mathcal{C})$ and $\mathrm{PL}(\mathcal{C})$ have decidable *membership*.

The next step is to tackle *separation*. This question is difficult in general, but it is worth looking at *particular* input classes. For instance, one can define the TL-hierarchy of basis $\mathcal{C}$: level 0 is $\mathrm{TL}_0(\mathcal{C}) = \mathcal{C}$ and level $n \geq 1$ is $\mathrm{TL}_n(\mathcal{C}) = \mathrm{TL}(\mathrm{TL}_{n-1}(\mathcal{C}))$. It can be shown that the hierarchies of bases $\mathrm{ST} = \{\emptyset, A^*\}$ and $\mathrm{DD} = \{\emptyset, \{\varepsilon\}, A^+, A^*\}$ are strict. Thus, since $\mathrm{BPol}(\mathcal{C}) \subseteq \mathrm{TL}(\mathcal{C})$, they both classify the star-free languages (or equivalently the languages definable in full linear temporal logic). We already know that in both hierarchies, membership is decidable for levels 1 (*i.e.*, the variants TL and TLX of unary temporal logic) and 2 (which were studied in [15]). The results of the present paper show that if $\mathrm{TL}_2(\mathrm{ST})$ and $\mathrm{TL}_2(\mathrm{DD})$ have decidable separation, then $\mathrm{TL}_3(\mathrm{ST})$ and $\mathrm{TL}_3(\mathrm{DD})$ would have decidable membership.

Finally, all other major operators have language-theoretic counterparts. Another possible follow-up is to look for such a definition for all three operators $\mathcal{C} \mapsto \mathrm{TL}(\mathcal{C}), \mathrm{FL}(\mathcal{C})$ and $\mathrm{PL}(\mathcal{C})$.

References

1 Mustapha Arfi. Polynomial operations on rational languages. In *Proceedings of the 4th Annual Symposium on Theoretical Aspects of Computer Science, STACS'87*, Lecture Notes in Computer Science, pages 198–206. Springer, 1987.

2 David A. Mix Barrington, Kevin Compton, Howard Straubing, and Denis Thérien. Regular languages in NC^1. *Journal of Computer and System Sciences*, 44(3):478–499, 1992.

3 Danièle Beauquier and Jean-Éric Pin. Languages and scanners. *Theoretical Computer Science*, 84(1):3–21, 1991.

4 Janusz A. Brzozowski and Imre Simon. Characterizations of locally testable events. *Discrete Mathematics*, 4(3):243–271, 1973.

5 Luc Dartois and Charles Paperman. Two-variable first order logic with modular predicates over words. In *Proceedings of the 30th International Symposium on Theoretical Aspects of Computer Science, STACS'13*, Leibniz International Proceedings in Informatics (LIPIcs), pages 329–340. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2013.

6 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.

7 Volker Diekert, Martin Horsch, and Manfred Kufleitner. On first-order fragments for Mazurkiewicz traces. *Fundamenta Informaticae*, 80(1-3):1–29, 2007.

8 Volker Diekert and Manfred Kufleitner. Fragments of first-order logic over infinite words. *Theory of Computing Systems (ToCS)*, 48(3):486–516, 2011.

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.

10 James Alexander Green. On the structure of semigroups. *Annals of Mathematics*, 54(1):163–172, 1951.

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.

12 Robert Knast. A semigroup characterization of dot-depth one languages. *RAIRO – Theoretical Informatics and Applications*, 17(4):321–330, 1983.

13 Andreas Krebs, Kamal Lodaya, Paritosh K. Pandya, and Howard Straubing. Two-variable logic with a between relation. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS'16*, pages 106–115, 2016.

14 Andreas Krebs, Kamal Lodaya, Paritosh K. Pandya, and Howard Straubing. An algebraic decision procedure for two-variable logic with a between relation. In *27th EACSL Annual Conference on Computer Science Logic, CSL'18*, Leibniz International Proceedings in Informatics (LIPIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018.

15 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*, Volume 16, Issue 3, 2020.

16 Robert McNaughton and Seymour A. Papert. *Counter-Free Automata.* MIT Press, 1971.

17 Jean-Éric Pin. *Varieties of Formal Languages.* North Oxford Academic, 1986.

18 Jean-Éric Pin and Pascal Weil. Polynomial closure and unambiguous product. *Theory of Computing Systems*, 30(4):383–422, 1997.

19 Jean-Éric Pin. An explicit formula for the intersection of two polynomials of regular languages. In *Proceedings of the 17th International Conference on Developments in Language Theory, DLT'13*, volume 7907 of *Lecture Notes in Computer Science*, pages 31–45. Springer, 2013.

20 Jean-Éric Pin. Mathematical foundations of automata theory. Lecture notes, in preparation, 2022. URL: `https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf`.

21 Thomas Place. The amazing mixed polynomial closure and its applications to two-variable first-order logic. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS'22*, 2022.

22 Thomas Place and Marc Zeitoun. Separating without any ambiguity. In *45th International Colloquium on Automata, Languages, and Programming, ICALP'18*, Leibniz International Proceedings in Informatics (LIPIcs), pages 137:1–137:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018.

23 Thomas Place and Marc Zeitoun. Generic results for concatenation hierarchies. *Theory of Computing Systems (ToCS)*, 63(4):849–901, 2019. Selected papers from CSR'17.

24 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.

25 Thomas Place and Marc Zeitoun. On all things star-free. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming, ICALP'19*, Leibniz International Proceedings in Informatics (LIPIcs), pages 126:1–126:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019.

26 Thomas Place and Marc Zeitoun. How many times do you need to go back to the future in unary temporal logic? In *Proceedings of the 15th Latin American Symposium on Theoretical Informatics, LATIN'22*, Lecture Notes in Computer Science. Springer, 2022.

27 Thomas Place and Marc Zeitoun. All about unambiguous polynomial closure. *TheoretiCS*, 2(11):1–74, 2023. `doi:10.46298/theoretics.23.11`.

28 Thomas Place and Marc Zeitoun. Closing star-free closure, 2023. `arXiv:2307.09376`.

29 Thomas Place and Marc Zeitoun. A generic characterization of generalized unary temporal logic and two-variable first-order logic, 2023. `arXiv:2307.09349`.

30 Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8(2):190–194, 1965.

31 Marcel Paul Schützenberger. Sur certaines opérations de fermeture dans les langages rationnels. *Symposia Mathematica*, XV:245–253, 1975.

32 Marcel Paul Schützenberger. Sur le produit de concaténation non ambigu. *Semigroup Forum*, 13:47–75, 1976.

33 Imre Simon. Piecewise testable events. In *Proceedings of the 2nd GI Conference on Automata Theory and Formal Languages*, pages 214–222. Springer, 1975.

34 Howard Straubing. Aperiodic homomorphisms and the concatenation product of recognizable sets. *Journal of Pure and Applied Algebra*, 15(3):319–327, 1979.

35 Pascal Tesson and Denis Thérien. Diamonds are forever: The variety DA. In *Semigroups, Algorithms, Automata and Languages*, pages 475–500. World Scientific, 2002.

36 Denis Thérien and Thomas Wilke. Over words, two variables are as powerful as one quantifier alternation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing, STOC'98*, pages 234–240. ACM, 1998.