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

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


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] ($FO(\prec)$) 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
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 \( C \) as input, and outputs a larger one \( Op(C) \). Thus, we can study all classes \( Op(C) \) simultaneously: “what hypotheses about \( C \) guarantee the decidability of \( Op(C) \)-membership?”. For example, one can generalize the three definitions of star-free languages through operators:

1. The star-free closure \( C \rightarrow \text{SF}(C) \) has been introduced in [31, 34]. Languages in \( \text{SF}(C) \) are defined by “extended” star-free expressions, which can freely use languages from \( C \).

2. A construction associating a signature \( I_C \) to a class \( C \) has been given in [23]. For each \( L \in C \), the set \( I_C \) 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 \( C \rightarrow \text{FO}(I_C) \) based on first-order logic. It captures many choices of signature.

3. Similarly, an operator \( C \rightarrow \text{LTL}(C) \) that generalizes LTL has been defined in [28]. It is shown in [23, 28] that \( \text{SF}(C) = \text{FO}(I_C) = \text{LTL}(C) \) for any class \( C \) (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 \( C \text{-orbits:} \quad L \in \text{SF}(C) \text{ if and only if its } C \text{-orbits are all aperiodic. This elegantly generalizes Schützenberger’s theorem and gives a transfer theorem for membership. Indeed, the } C \text{-orbits are connected with a decision problem that strengthens membership: } C \text{-separation. Given two input regular languages } L_1 \text{ and } L_2, \text{ } C \text{-separation asks whether there is } K \in C \text{ such that } L_1 \subseteq K \text{ and } L_2 \cap K = \emptyset. \text{ The crucial point is that } C \text{-orbits are computable if } C \text{-separation is decidable. Thus, } \text{SF}(C) \text{-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 (\( \text{FO}^2(<) \)) 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 \( \text{FO}^2(<) = \text{TL} \). Its algebraic characterization by Thérien and Wilke [36] is one of the famous results of this type: a regular language belongs to \( \text{FO}^2(<) = \text{TL} \) if and only if its syntactic monoid belongs to the variety of monoids DA.

Both definitions extend to natural operators. First, the generic signatures \( I_C \) yield an operator \( C \rightarrow \text{FO}^2(I_C) \) based on \( \text{FO}^2 \). Second, an operator \( C \rightarrow \text{TL}(C) \) has been defined in [27]. It enriches TL with new modalities \( F_L \) and \( P_L \), both depending on the languages \( L \in C \). For example, the formula \( F_L \varphi \) holds at a position \( i \) in a word \( w \) if there is a position \( j > i \) in \( w \) such that \( \varphi \) holds at \( j \) and the infix between \( i \) and \( j \) belongs to \( L \). We know that \( \text{FO}^2(I_C) = \text{TL}(C) \) when \( C \) is closed under Boolean operations [27]. Here, we work with the \( \text{TL}(\cdot) \) operator, which encompasses all classic classes based on two-variable first-order logic or unary temporal logic. This includes the original variants \( \text{FO}^2(<) = \text{TL} \) and \( \text{FO}^2(<, +1) = \text{TLX} \), 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
We prove a generic algebraic characterization of the classes \( \text{FL}^2(\langle, \text{MOD} \rangle) = \text{TL}(C) \). We reuse the \( C \)-orbits introduced for star-free closure: for any class \( C \) (having mild properties) we show that a regular language belongs to \( \text{TL}(C) \) if and only if all \( C \)-orbits of its syntactic monoid belong to \( \text{DA} \). In particular, this yields a transfer theorem for membership: if \( C \) has decidable separation, then \( \text{FL}^2(\langle \rangle) = \text{TL}(C) \) 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 (\( \text{BPol} \)) and unambiguous polynomial closure (\( \text{UPol} \)). We have \( \text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C) \) if \( C \) has mild properties [27]. In fact, for many natural classes, \( \text{UPol}(\text{BPol}(C)) = \text{TL}(C) \). For example, \( \text{UPol}(\text{BPol}(\langle \emptyset, A^* \rangle)) \) is the class \( UL = \text{TL} = \text{FO}^2(\langle \rangle) \).

More generally, \( \text{UPol}(\text{BPol}(C)) = \text{TL}(C) \) for every class \( C \) consisting of group languages [27]. Yet, this is a strong hypothesis and the inclusion \( \text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C) \) is strict in general. For example, the results of [15] imply that \( \text{UPol}(\text{BPol}(\text{AT})) \neq \text{TL}(\text{AT}) \), where \( \text{AT} \) consists of all Boolean combinations of languages \( A^*aA^* \) (with \( a \in A \)). Nevertheless, the classes \( \text{UPol}(\text{BPol}(C)) \) serve as a central ingredient in the most difficult direction of our proof: “If a language satisfies our characterization on \( C \)-orbits, prove that it belongs to \( \text{TL}(C) \)”.

More precisely, we exploit the known characterization of \( \text{UPol}(\text{BPol}(C)) \) to prove that auxiliary languages belong to this class, and we then conclude using the inclusion \( \text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C) \).

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

**Contributions.** We prove a generic algebraic characterization of the classes \( \text{FO}^2(\langle, \text{MOD} \rangle) = \text{TL}(C) \). We reuse the \( C \)-orbits introduced for star-free closure: for any class \( C \) (having mild properties) we show that a regular language belongs to \( \text{TL}(C) \) if and only if all \( C \)-orbits of its syntactic monoid belong to \( \text{DA} \). In particular, this yields a transfer theorem for membership: if \( C \) has decidable separation, then \( \text{FO}^2(\langle \rangle) = \text{TL}(C) \) 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 (\( \text{BPol} \)) and unambiguous polynomial closure (\( \text{UPol} \)). We have \( \text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C) \) if \( C \) has mild properties [27]. In fact, for many natural classes, \( \text{UPol}(\text{BPol}(C)) = \text{TL}(C) \). For example, \( \text{UPol}(\text{BPol}(\langle \emptyset, A^* \rangle)) \) is the class \( UL = \text{TL} = \text{FO}^2(\langle \rangle) \).

More generally, \( \text{UPol}(\text{BPol}(C)) = \text{TL}(C) \) for every class \( C \) consisting of group languages [27]. Yet, this is a strong hypothesis and the inclusion \( \text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C) \) is strict in general. For example, the results of [15] imply that \( \text{UPol}(\text{BPol}(\text{AT})) \neq \text{TL}(\text{AT}) \), where \( \text{AT} \) consists of all Boolean combinations of languages \( A^*aA^* \) (with \( a \in A \)). Nevertheless, the classes \( \text{UPol}(\text{BPol}(C)) \) serve as a central ingredient in the most difficult direction of our proof: “If a language satisfies our characterization on \( C \)-orbits, prove that it belongs to \( \text{TL}(C) \)”.

More precisely, we exploit the known characterization of \( \text{UPol}(\text{BPol}(C)) \) to prove that auxiliary languages belong to this class, and we then conclude using the inclusion \( \text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C) \).

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

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

## 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 \( \varepsilon \). 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 \( \langle K, L \rangle \subseteq A^* \), 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 \( \text{Pos}(w) = \{0, 1, \ldots, |w|, |w| + 1\} \) of \( |w| + 2 \) positions. In addition, we let \( \text{Pos}_i(w) = \{1, \ldots, |w|\} \subseteq \text{Pos}(w) \). Position \( i \in \text{Pos}_i(w) \) 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 \( x, y, z \in A^* \) such that \( w = xvy \) (resp. \( w = yv, w = xv \)). Given a word \( w = a_1 \cdots a_{|w|} \in A^* \) and \( i, j \in \text{Pos}(w) \), such that \( i < j \), we write \( w(i, j) = a_i \cdots a_{j-1} \in A^* \) (i.e., the infix obtained by keeping the letters carried by positions strictly between \( i \) and \( j \)).
Classes. A class of languages $C$ is simply a set of languages. Such a class $C$ is a lattice when $\emptyset \in C$, $A^* \in C$ and $C$ is closed under both union and intersection: for all $K, L \in C$, we have $K \cup L \in C$ and $K \cap L \in C$. Moreover, a class of languages $C$ is a Boolean algebra if it is a lattice closed under complement: for all $L \in C$, we have $A^* \setminus L \in C$. Finally, the class $C$ is closed under quotients if for all $L \in C$ and $u \in A^*$, we have $u^{-1}L \overset{\text{def}}{=} \{w \in A^* \mid uw \in L\} \in C$ and $Lu^{-1} \overset{\text{def}}{=} \{w \in A^* \midwu \in L\} \in C$. 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 $\omega$ when $M$ is understood) such that for every $s \in M$, the element $s^\omega$ 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:

\[
\begin{align*}
\small{s} &<_\delta t \quad \text{when there exist } x, y \in M \text{ such that } s = xty, \\
\small{s} &\leq_L t \quad \text{when there exists } x \in M \text{ such that } s = xt, \\
\small{s} &\leq_R t \quad \text{when there exists } y \in M \text{ such that } s = ty.
\end{align*}
\]

Clearly, $\leq_\delta$, $\leq_L$ and $\leq_R$ are preorders (i.e., they are reflexive and transitive). We write $<_\delta$, $<_L$ and $<_R$ for their strict variants (for example, $s <_\delta t$ when $s <_\delta t$ but $t \not<_\delta s$). Finally, we write $\delta$, $\mathcal{L}$ and $\mathcal{R}$ for the corresponding equivalence relations (for example, $s \delta t$ when $s <_\delta t$ and $t <_\delta s$). 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]).

\textbf{Lemma 1.} Let $M$ be a finite monoid and let $s, t \in M$. If $s \delta t$ and $s \leq_R t$, then $s \mathcal{R} t$.

Regular languages and syntactic morphisms. Since $A^*$ is a monoid whose multiplication is concatenation (the identity element is $\varepsilon$), 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 $\alpha$ 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 $xuy \in L \iff xyv \in L$ 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 $C$. They serve as mathematical tools for analyzing it, as obtaining an algorithm for one of these problems requires a solid understanding of that class $C$. The $C$-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 \( L_2 \).

\[ \text{Lemma 4.} \] 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 \( \mathcal{C} \) operators. These are correspondences \( \mathcal{C} 
\mapsto \operatorname{Op}(\mathcal{C}) \) that take as input a class \( \mathcal{C} \) to build a new one \( \operatorname{Op}(\mathcal{C}) \). 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 \( \alpha : A^* \rightarrow M \), we define special subsets of \( M \): the \( \mathcal{C} \)-orbits for \( \alpha \).

This notion was introduced in [28]. We shall use it to formulate generic characterizations of the operators \( \mathcal{C} 
\mapsto \operatorname{Op}(\mathcal{C}) \) that we consider: for each input prevariety \( \mathcal{C} \), the languages in \( \operatorname{Op}(\mathcal{C}) \) are characterized by a property of the \( \mathcal{C} \)-orbits for their syntactic morphisms.

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

\[ \text{Lemma 3.} \] Let \( \mathcal{C} \) be a prevariety and \( \alpha : A^* \rightarrow M \) be a morphism. Then,

1. For every \( \mathcal{C} \)-morphism \( \eta : A^* \rightarrow N \), all \( \mathcal{C} \)-pairs for \( \alpha \) are also \( \eta \)-pairs for \( \alpha \).
2. There exists a \( \mathcal{C} \)-morphism \( \eta : A^* \rightarrow N \) such that all \( \eta \)-pairs for \( \alpha \) are also \( \mathcal{C} \)-pairs for \( \alpha \).

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

\[ \text{Lemma 4.} \] Let \( \mathcal{C} \) be a prevariety and \( \alpha : A^* \rightarrow M \) be a surjective morphism into a finite monoid. For all \( e \in E(M) \), the \( \mathcal{C} \)-orbit of \( e \) for \( \alpha \) 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.

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

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 $\alpha$ is contained in the set $\alpha(\eta^{-1}(f))$.

Proof. Let $t_1, \ldots, t_n \in M$ be all elements of the set $\{t \in M \mid (e, t) \text{ is a C-pair}\}$. By definition, the $\mathcal{C}$-orbit of $e$ for $\alpha$ is $M_e = \{\eta(t_1)e, \ldots, \eta(t_n)e\}$. Since $\eta$ is a $\mathcal{C}$-morphism, Lemma 3 implies that $(e, t_i)$ is an $\eta$-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 $et_i \in \alpha(\eta^{-1}(f))$ for $i \leq n$. We define $w_i = (x_1 \cdots x_n)^px_1 \cdots x_{i-1}y_ix_{i+1} \cdots x_n(x_1 \cdots x_n)^{2p-1}$. By definition, we have $\alpha(w_i) = et_i$. Now, since $\eta(x_i) = \eta(y_i)$, we get $\eta(w_i) = \eta(w) = f$. Hence, $et_i \in \alpha(\eta^{-1}(f))$, as desired. ▶

4 Generalized unary temporal logic

In this section, we define generalized unary temporal logic. We introduce an operator $\mathcal{C} \mapsto \text{TL}(\mathcal{C})$ that associates a new class of languages $\text{TL}(\mathcal{C})$ 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 $\text{TL}(\mathcal{C})$ as follows. A $\text{TL}(\mathcal{C})$ formula is built from atomic formulas using Boolean connectives and temporal operators. The atomic formulas are $\top$, $\bot$, $\text{min}$, $\text{max}$ and “$a$” for every letter $a \in A$. All Boolean connectives are allowed: if $\psi_1$ and $\psi_2$ are $\text{TL}(\mathcal{C})$ formulas, then so are $(\psi_1 \lor \psi_2)$, $(\psi_1 \land \psi_2)$ 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 $\psi$ is a $\text{TL}(\mathcal{C})$ formula, then so are $(F_L \psi)$ and $(P_L \psi)$. For the sake of improved readability, we omit parentheses when there is no ambiguity.

Semantics. Evaluating a $\text{TL}(\mathcal{C})$ formula $\varphi$ requires a word $w \in A^*$ and a position $i \in \text{Pos}(w)$. We define by induction what it means for $(w, i)$ to satisfy $\varphi$, 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 $\ell \in A \cup \{\text{min}, \text{max}\}$, $w, i \models \ell$ holds when $\ell = w[i]$.

- Disjunction: $w, i \models \psi_1 \lor \psi_2$ when $w, i \models \psi_1$ or $w, i \models \psi_2$.

- Conjunction: $w, i \models \psi_1 \land \psi_2$ 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 $w, i \models F_L \psi$ when there exists $j \in \text{Pos}(w)$ such that $i < j$, $w(i, j) \in L$ and $w, j \models \psi$.

- Previously: for $L \in \mathcal{C}$, we let $w, i \models P_L \psi$ when there exists $j \in \text{Pos}(w)$ 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 $\text{TL}(\mathcal{C})$ for which this choice does matter. This is why we introduce notations for both conventions. Given a formula $\varphi \in \text{TL}(\mathcal{C})$ we let $L_{\text{min}}(\varphi) = \{w \in A^* \mid w, 0 \models \varphi\}$ and $L_{\text{max}}(\varphi) = \{w \in A^* \mid w, |w| + 1 \models \varphi\}$.

We are now ready to define the operator $\mathcal{C} \mapsto \text{TL}(\mathcal{C})$. Consider an arbitrary class $\mathcal{C}$. We write $\text{TL}(\mathcal{C})$ for the class consisting of all languages $L_{\text{min}}(\varphi)$ where $\varphi \in \text{TL}(\mathcal{C})$. Observe that by definition, $\text{TL}(\mathcal{C})$ is a Boolean algebra. Actually, the results of [27] imply that when $\mathcal{C}$ is a prevariety, then so is $\text{TL}(\mathcal{C})$ (we do not need this fact in the present paper).
**Classic unary temporal logic.** Let $ST = \{0, A^*\}$ and $DD = \{0, \{\epsilon\}, A^+, A^*\}$. The modalities $F_{A^*}$ and $P_{A^*}$ have the same semantics as the modalities $F$ and $P$ of standard unary temporal logic – e.g., $w, i \models F \varphi$ when there exists $j \in \text{Pos}(w)$ such that $i < j$ and $w, j \models \varphi$. Similarly, the modalities $F_{\{\epsilon\}}$ and $P_{\{\epsilon\}}$ have the same semantics as the modalities $X$ (next) and $Y$ (yesterday) – e.g., $w, i \models X \varphi$ when $i + 1 \in \text{Pos}(w)$ 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 $TL = TL(ST)$ and $TLX = TL(DD)$.

> **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 $TL(DD \setminus \{A^+\})$ are identical. However, in order to use generic results such as those from Section 3, we require the classes to which the operator $C \mapsto TL(C)$ 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 $FO^2(<)$ (two-variable first-order logic equipped with the linear order), and that $TLX$ corresponds to $FO^2(<, +1)$ (which also allows the successor). In [27], these results are generalized to all classes $TL(C)$ where $C$ is a Boolean algebra. In this case, we can construct from $C$ a set of predicates $\ll_{\text{C}}$ such that $TL(C) = FO^2(\ll_{\text{C}})$.

> **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 $FO^2(\ll_{\text{AT}})$, 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 $TL(C)$.

### 4.2 Connection with unambiguous polynomial closure

It is shown in [27] that $C \mapsto TL(C)$ can be expressed by other operators for very specific inputs: prevarieties of group languages. If $G$ is such a class, then $TL(G)$ coincides with $UPol(BPol(G))$, a class built on top of $G$ 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, $UPol(BPol(C))$ is strictly included in $TL(C)$ (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 $L_0, \ldots, L_n \subseteq A^*$, a **marked product** of $L_0, \ldots, L_n$ is a product of the form $L_0a_1L_1 \cdots a_nL_n$ where $a_1, \ldots, a_n \in A$. A single language $L_0$ is a marked product (this is the case $n = 0$). The **polynomial closure** of a class $C$, denoted by $Pol(C)$, consists of all finite unions of marked products $L_0a_1L_1 \cdots a_nL_n$ such that $L_0, \ldots, L_n \in C$. If $C$ is a prevariety, then $Pol(C)$ is a lattice (this is due to Arfi [1], see also [19, 23] for recent proofs). However, $Pol(C)$ need not be closed under complement. This is why it is often combined with another operator: the Boolean closure of a class $D$, denoted by $Bool(D)$, is the least Boolean algebra containing $D$. We write $BPol(C)$ for $Bool(Pol(C))$. It is standard that if $C$ is a prevariety, then so is $BPol(C)$ (see [23] for example). Finally, $UPol$ is the unambiguous restriction of $Pol$. A marked product $L_0a_1L_1 \cdots a_nL_n$ is unambiguous when every word $w \in L_0a_1L_1 \cdots a_nL_n$ has a unique decomposition $w = w_0a_1w_1 \cdots a_nw_n$ where $w_i \in L_i$ for $0 \leq i \leq n$. The **unambiguous polynomial closure** of a class $C$, written $UPol(C)$, consists of all finite disjoint unions of unambiguous marked products $L_0a_1L_1 \cdots a_nL_n$ such that $L_0, \ldots, L_n \in C$ (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 $C$ is a prevariety, then so is $UPol(C)$. Thus, $UPol$ preserves closure under complement.
In the paper, we are interested in the “combined” operator $C \mapsto \text{UPol}(\text{BPol}(C))$. Indeed, it is connected to the classes $\text{TL}(C)$ by the following proposition proved in [27, Proposition 9.12].

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

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

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

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

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

We present a generic characterization of $\text{TL}(C)$ when $C$ 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 $\text{TL}(\text{AT}) = \text{FO}^2(\text{I}_{\text{AT}})$ by Krebs, Lodaya, Pandya and Straubing [13, 14, 15].

### 5.1 Statement

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

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

Thérien and Wilke [36] showed that a regular language belongs to $\text{TL}$ if and only if its syntactic monoid is in $\text{DA}$ (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 $\text{TL}(C)$, based on $C$-orbits introduced in Section 3.

**Theorem 12.** Let $C$ 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. $L \in \text{TL}(C)$.
2. For every idempotent $e \in E(M)$, the $C$-orbit of $e$ for $\alpha$ belongs to $\text{DA}$.

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

**Corollary 13.** If a prevariety $C$ has decidable separation, $\text{TL}(C)$ 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 $C$-orbit of $e \in E(M)$ for $\alpha$ belongs to $\text{DA}$ means that we have,

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

One can check that (3) follows from (1), which characterizes $\text{UPol}(\text{BPol}(C))$ (this is consistent with Proposition 10 asserting that $\text{UPol}(\text{BPol}(C)) \subseteq \text{TL}(C)$). Indeed, choosing $t = s$ in (1) shows that the $C$-orbit of $e$ is aperiodic, i.e., $(esete)^{\omega+1} = (eset)^{\omega}$ if $(e, s)$ is a $C$-pair. However, note that the element $t$ is “free” in (1), whereas it must be part of a $C$-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(\text{AT}), as mentioned at the beginning of the section.
5.2 Application to historical classes

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

\[
\text{Theorem 15 (Thérien and Wilke [36]). Let } L \subseteq A^* \text{ be a regular language and let } M \text{ be its syntactic monoid. The two following properties are equivalent:}
\]

1. \( L \) belongs to TL.
2. \( M \) belongs to DA.

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

We turn to the characterization of TLX = TL(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 \( e \).

\[
\text{Theorem 16 (Thérien and Wilke [36]). Let } L \subseteq A^* \text{ be a regular language and } S \text{ be its syntactic semigroup. The two following properties are equivalent:}
\]

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 \( \alpha \). Since 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 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 \( M_{1M} = \{ 1_M \} \) (which clearly belongs to DA). Hence, every DD-orbit for \( \alpha \) belongs to DA if and only if \( eSe \in DA \) for every \( e \in E(S) \). In view of Theorem 12, this implies Theorem 16. \( \square \)

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 \( M_eDA \). 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 \( \{ s \in M | e \leq_s s \} \). We say that \( M \) belongs to \( M_eDA \) if and only if for every idempotent \( e \in E(M) \), the monoid of \( eN_e \) belongs to DA.

\[
\text{Theorem 17 (Krebs, Lodaya, Pandya and Straubing [15]). Let } L \subseteq A^* \text{ be a regular language and } M \text{ be its syntactic monoid. The two following properties are equivalent:}
\]

1. \( L \in TL(AT) \).
2. \( M \) belongs to \( M_eDA \).

**Proof.** For \( w \in A^* \), let \( \text{alph}(w) \subseteq A \) 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 \( \alpha \). We prove that \( M_e = eN_e \) for every \( e \in E(M) \). It will follows that \( M \) belongs to \( M_eDA \) if and only if every AT-orbit for \( \alpha \) belongs to DA. In view of Theorem 12 this implies Theorem 17.

We first consider \( s' \in eN_e \) 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 \( e \leq_{\leq} s_i \) 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 \( e \leq_{\leq} s_i \), we have \( q_isr_i \in M \) such that \( e = q_isr_i \) for every \( i \leq n \). Hence, since \( e \in E(M) \), we have \( e = q_isr_1 \cdots q_isnr_n \). For every
i ≤ n, let \( x_i \in \alpha^{-1}(y_i), y_i \in \alpha^{-1}(r_i) \) 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 \( \text{alph}(w) = \text{alph}(w') \). 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 \( \alpha \), 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 \( \text{alph}(u) = \text{alph}(v) \), \( \alpha(u) = e \) and \( \alpha(v) = s \). Let \( a_1 \ldots , a_n \in A \) be the letters such that \( v = a_1 \cdots a_n \). Since \( \text{alph}(u) = \text{alph}(v) \), it is immediate that for each \( i \leq n \), there are \( x_i, y_i \in A^* \) such that \( u = x_i a_i y_i \). Hence \( e = \alpha(u) \leq_\beta \alpha(a_i) \) and we conclude that \( s = \alpha(a_1 \cdots a_n) \in N_e \). Consequently, \( s' = ese \in eN_e e \), 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 \( L \in \text{TL}(\mathcal{C}) \) if and only if all \( \mathcal{C} \)-orbits for \( \alpha \) belong to \( \text{DA} \). We start with the left-to-right implication.

From \( \text{TL}(\mathcal{C}) \) to \( \text{DA} \). This direction follows from results of [27]. To use them, we need some preliminary terminology. We introduce equivalence relations connected to the class \( \text{TL}(\mathcal{C}) \) 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 \( \eta \). The following fact is easy (see [27, Fact 9.3]).

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

We use the standard notion of rank of a \( \text{TL}[\mathcal{C}_\eta] \) formula: the rank of \( \varphi \) 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 \( \varphi \).
- The rank of \( \varphi \land \psi \) and \( \varphi \lor \psi \) is the maximum between the ranks of \( \varphi \) and \( \psi \).
- For every language \( L \subseteq A^* \), the rank of \( F_L \varphi \) and \( P_L \varphi \) is the rank of \( \varphi \) plus 1.

Two \( \text{TL}[\mathcal{C}_\eta] \) formulas \( \varphi \) and \( \psi \) are equivalent if they have the same semantics. That is, for every \( w \in A^* \) and every position \( i \in \text{Pos}(w) \), we have \( w, i \models \varphi \iff w, i \models \psi \). 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 \( \text{TL}[\mathcal{C}_\eta] \) 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 \( w, w' \in A^* \), \( i \in \text{Pos}(w) \) and \( i' \in \text{Pos}(w') \), we write, \( w, i \equiv_{\eta,k} w', i' \) when:

For every \( \text{TL}[\mathcal{C}_\eta] \) formula \( \varphi \) of rank at most \( k \), \( w, i \models \varphi \iff w', i' \models \varphi \).

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

Lemma 20. Let \( \mathcal{C} \) be a prevariety and \( L \subseteq A^* \). If \( L \in \text{TL}(\mathcal{C}) \), then there exists a \( \mathcal{C} \)-morphism \( \eta : A^* \to N \) and \( k \in \mathbb{N} \) such that \( L \) is a union of \( \equiv_{\eta,k} \)-classes.
Proof. Let \( L \in \text{TL}(C) \). There exists a \( \text{TL}[C] \) formula \( \varphi \) such that \( w \in L \iff w, 0 \models \varphi \) for all \( w \in A^* \). By Fact 18, there exists a \( C \)-morphism \( \eta : A^* \to N \) such that \( \varphi \) is a \( \text{TL}[C^r] \) formula. Let \( k \in \mathbb{N} \) be the rank of \( \varphi \). We prove that \( L \) is a union of \( \equiv_{\omega,k} \)-classes. Given \( w, w' \in A^* \) such that \( w \equiv_{\omega,k} w' \), we have to prove that \( w \in L \iff w' \in L \). By symmetry, we only prove the left to right implication. Thus, we assume that \( w \in L \). By definition of \( \varphi \), it follows that \( w, 0 \models \varphi \). Moreover, since \( w \equiv_{\omega,k} w' \) (i.e., \( w, 0 \equiv_{\omega,k} w', 0 \)) and \( \varphi \) is a \( \text{TL}[C^r] \) formula of rank \( k \), we have \( w, 0 \models \varphi \) by definition of \( \equiv_{\omega,k} \). Hence, \( w' \in L \) by definition of \( \varphi \), as desired. \( \blacklozenge \)

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

\begin{proposition}
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:

\[
x(z^kuz^{2k}vz^k)^k(z^kuz^{2k}vz^k)^k y \equiv_{\omega,k} x(z^kuz^{2k}vz^k)^k z^k vz^k (z^kuz^{2k}vz^k)^k y.
\]

We are ready to conclude this direction of the proof: assuming that \( L \in \text{TL}(C) \), we show that all \( C \)-orbits for its syntactic monoid belong to DA. Let \( e \in E(M) \) and \( M_e \) be its \( C \)-orbit. Proving that \( M_e \in \text{DA} \) amounts to proving that any elements \( s, t \in M_e \) satisfy (2). Fix \( s, t \in E(M) \times M_e \times M_e \). Lemma 20 yields a \( C \)-morphism \( \eta : A^* \to N \) and \( k \in \mathbb{N} \) such that \( L \) is a union of \( \equiv_{\omega,k} \)-classes. Since \( \eta \) is a \( C \)-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,

\[
x(z^kuz^{2k}vz^k)^k(z^kuz^{2k}vz^k)^k y \equiv_{\omega,k} x(z^kuz^{2k}vz^k)^k z^k vz^k (z^kuz^{2k}vz^k)^k y.
\]

Since \( L \) is a union of \( \equiv_{\omega,k} \)-classes, the words \( (z^kuz^{2k}vz^k)^k \) and \( (z^kuz^{2k}vz^k)^k z^k vz^k (z^kuz^{2k}vz^k)^k \) are equivalent for the syntactic congruence of \( L \), so they have the same image under its syntactic morphism \( \alpha \). 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 \( (st)^{\omega} = (st)^k t(st)^{-k} \). Therefore, (2) holds.

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

Since \( \text{TL}(C) \) is closed under union, it suffices to prove that for all \( s \in M \), the language \( \alpha^{-1}(s) \) is in \( \text{TL}(C) \). We achieve this by inductively constructing a \( \text{TL}[C] \) formula defining \( \alpha^{-1}(s) \). According to Lemma 3, there exists a \( C \)-morphism \( \eta : A^* \to N \) such that the \( C \)-pairs for \( \alpha \) are exactly the \( \eta \)-pairs for \( \alpha \). We use \( \eta \) to leverage the assumption that all \( C \)-orbits for \( \alpha \) belong to DA. More precisely, \( \eta \) recognizes all the basic languages in \( C \) that we shall use in our \( \text{TL}[C] \) formulas. The induction proceeds as follows: using \( \eta \), we define a sequence of languages \( K_0 \supseteq K_1 \supseteq \cdots \supseteq K_{|N|} \) and show by induction on \( |N| - t \) that \( K_{|N|} \cap \alpha^{-1}(t) \) can be defined by a \( \text{TL}[C] \) formula for each \( t \leq |N| \). The induction basis is the case \( t = |N| \), which is simple because \( K_{|N|} \) is a finite language. Furthermore, the case \( t = 0 \) gives the desired result since \( K_0 \) contains all words. The induction step consists in building a \( \text{TL}[C] \) formula describing \( K_{t+1} \cap \alpha^{-1}(t) \) from several \( \text{TL}[C] \) formulas that describe the languages \( K_{t+1} \cap \alpha^{-1}(t) \) 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 \( K_{t+1} \cap \alpha^{-1}(t) \) 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 \( K_{t+1} \cap \alpha^{-1}(s) \), 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 $\alpha$ and $\eta$. This is problematic, because such a letter does not necessarily exist. We solve this issue by considering an extended alphabet $B$, replacing $\alpha : A^* \rightarrow M$ and $\eta : A^* \rightarrow N$ with two new morphisms $\beta : B^* \rightarrow M$ and $\delta : B^* \rightarrow N$ 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 $\mathsf{TL}(C)$) and our hypothesis (that every $C$-orbit for $\alpha$ belongs to $\mathsf{DA}$) on the new morphisms $\beta$ and $\eta$.

We now start the proof by first defining $\beta$ and $\delta$. Recall that $\eta : A^* \rightarrow N$ is the $C$-morphism provided by Lemma 3: it is such that the $C$-pairs for $\alpha$ are exactly the $\eta$-pairs for $\alpha$. We fix $\eta$ 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 $\beta : B^* \rightarrow M$ and $\delta : B^* \rightarrow N$ 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 $C$ be the class of all languages over $B$ recognized by $\delta$. One can check that $C_\delta$ is a prevariety. We now reduce membership of inverse images under $\alpha$ to $\mathsf{TL}(C)$ to that of inverse images under $\beta$ to $\mathsf{TL}(C_\delta)$.

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

**Proof.** We first define a morphism $\gamma : A^* \rightarrow B^*$. 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 $K \in \mathsf{TL}(C_\delta)$, we have $\gamma^{-1}(K) \in \mathsf{TL}(C)$. We fix $K$ for the proof. Since $K \in \mathsf{TL}(C_\delta)$, it is defined by a formula $\psi \in \mathsf{TL}[C_\delta]$. We apply two kinds of modifications to $\psi$ in order to build a new formula $\psi' \in \mathsf{TL}[C]$ defining $\gamma^{-1}(K)$:

1. We replace every atomic subformula "$b" for $b \in B$ by the $\mathsf{TL}[C]$-formula $\bigvee_{a \in A} \psi(\alpha(a), \eta(a)) = a$.

2. For every temporal modality $F_H$ (resp. $P_H$) occurring in $\psi$, we have $H \in C_\delta$ by hypothesis.

Hence, $H$ is recognized by $\delta$ and there exists $G \subseteq N$ such that $H = \delta^{-1}(G)$. Note that $\eta^{-1}(G) \in C$ since $\eta$ is a $C$-morphism. We replace the temporal modality $F_H$ (resp. $P_H$) by $F_{\eta^{-1}(G)}$ (resp. $P_{\eta^{-1}(G)}$).

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

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

**Lemma 23.** For every $c \in E(M)$ and every $s, t \in M$, if $(c, s)$ and $(c, t)$ are $\delta$-pairs for $\beta$, then $(cse)c = (cse)c (cse)c$.

**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 $\beta$ and $\delta$ 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, $(c, s) \in M^2$ and $(c, t) \in M^2$ are $\eta$-pairs for $\alpha$. By definition of $\eta$, it follows that they are $C$-pairs for $\alpha$. Hence, $cse$ and $cet$ both belong to the $C$-orbit of $c$ for $\alpha$. Since all $C$-orbits for $\alpha$ belong to $\mathsf{DA}$ by hypothesis, this gives $(cse)c = (cse)c (cse)c$. $\blacksquare$
We now use the Green relation $\mathcal{J}$ over $N$ to associate a number $d_3(r) \in \mathbb{N}$ with every element $r \in N$. We let $d_2(r)$ be the maximal number $n \in \mathbb{N}$ such that there exist $n$ elements $r_1, \ldots, r_n \in N$ satisfying $r <_3 r_1 <_3 \cdots <_3 r_n$. By definition, $0 \leq d_2(r) \leq |N| - 1$. In particular, we have $d_2(r) = 0$ if and only if $r$ is maximal for $\leq_3$ (i.e., if and only if $r \not\mathcal{J} 1_N$).

Finally, given a word $w \in B^*$, we write $d_3(w) \in \mathbb{N}$ for $d_3(\delta(w))$. Observe that for all $x, y, z \in B^*$, we have $d_3(y) \leq d_3(xyz)$ (as $xyz \leq_3 y$), 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:

$$K_\ell = \{w \in B^* \mid \text{for all } k \leq \ell \text{ and } x, y, z \in B^*, \text{ if } w = xyz \text{ and } |y| = k, \text{ then } d_3(y) \geq k\}.$$ 

Note that $K_0 = B^*$ as $d_3(y) \geq 0$ for all $y \in B^*$. Also, if $\ell \geq |N|$, then $K_\ell$ is finite (it contains words of length at most $|N| - 1$ as $d_3(y) < |N|$ for all $y \in B^*$). We now have the next lemma.

**Lemma 24.** Let $\ell \in \mathbb{N}$ and $w \in K_\ell$. Then $d_3(w) \leq \ell$ if and only if for all $x, y, z \in B^*$ such that $w = xyz$ and $|y| \leq \ell + 1$, we have $d_3(y) \leq \ell$.

**Proof.** The “only if” direction is immediate since $d_3(y) \leq d_3(w)$ 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 $d_3(y) \leq \ell$. We prove that $d_3(w) \leq \ell$. If $|w| \leq \ell + 1$, this is immediate. Assume now that $|w| > \ell + 1$. We get $b_1, \ldots, b_n \in B$ and $v \in B^*$ such that $|v| = \ell + 1$ and $w = vb_1 \cdots b_n$. We use induction on $i$ to prove that $\delta(v) \not\mathcal{J} \delta(vb_1 \cdots b_i)$ for all $i \leq n$. Since $d_3(v) \leq \ell$ by hypothesis, the case $i = n$ yields $d_3(w) \leq \ell$. The case $i = 0$ is trivial: we have $\delta(v) \not\mathcal{J} \delta(v)$. Assume now that $i \geq 1$. By induction hypothesis, we know that $\delta(v) \not\mathcal{J} \delta(vb_1 \cdots b_{i-1})$. 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 $d_3(y) \geq \ell$. Moreover, $yb_i$ is an infix of $w$ such that $|yb_i| = \ell + 1$, which yields $d_3(yb_i) \leq \ell$ by hypothesis. Since $d_3(y) \leq d_3(yb_i)$, we get $d_3(yb_i) = d_3(y) = \ell$, which implies that $\delta(yb_i) \not\mathcal{J} \delta(y)$. Moreover, we have $\delta(yb_i) \leq \delta(y)$. Thus, Lemma 1 yields $\delta(yb_i) \not\mathcal{J} \delta(y)$. This implies that $\delta(xy) = \delta(yb_i) \not\mathcal{J} \delta(xy)$. Hence, $\delta(vb_1 \cdots b_i) \not\mathcal{J} \delta(vb_1 \cdots b_{i-1}) \not\mathcal{J} \delta(v)$. This completes the proof.

We now prove that for all $s \in M$ and $\ell \in \mathbb{N}$, we have $K_\ell \cap \beta^{-1}(s) \subseteq TL(C_3)$. Our objective (every language recognized by $\beta$ belongs to $TL(C_3)$) 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 $\delta$ has a $d_3$ value at most $\ell$. We do not use induction for this case, which relies on the inclusion $UPol(BPol(C_3)) \subseteq TL(C_3)$. It is also the place where we use Lemma 23, i.e., the hypothesis that all $C$-orbits for $\alpha$ are in DA.

**Proposition 25.** Let $(\ell, s, r) \in \mathbb{N} \times M \times N$. If $d_3(r) \leq \ell$ then $K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r) \subseteq TL(C_3)$.

**Proof.** We prove that $K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r) \subseteq UPol(BPol(C_3))$, which, by Proposition 10, will give the desired result $K_\ell \cap \beta^{-1}(s) \cap \delta^{-1}(r) \subseteq TL(C_3)$. 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 $C_3$-pair for $\gamma$, the following equation holds:

$$((f_{q_1}f_{q_2})^+) = (f_{q_1}f_{q_2})^+f_{q_2}f_{(f_{q_1}f_{q_2})^+}.$$ (4)

Let $q_1, q_2, f \in Q$ be such elements. By definition of $C_3$, we know that $\delta$ is a $C_3$-morphism. Therefore, Lemma 3 implies that $(f, q_1)$ is a $\delta$-pair for $\gamma$. We get $u', v' \in B^*$, such that $\delta(u') = \delta(v')$, $\delta(u') = f$ and $\gamma(v') = q_1$. Note that if $v' = 1$, then $q_1 = 1_Q$ and (4) holds since it is clear that $(f_{q_1}f)_{q_2}f_{+} = (f_{q_1}f)^{2+1}$. Therefore, we assume from now on that $v' \in B^*$. Let us also choose $v' \in B^*$ such that $\gamma(v') = q_2$. We now define $p = \ell \times \omega(N) \times \omega(M) \times \omega(Q)$,
\[ u = (u')^p, v_1 = (u')^{p-1}v_1' \text{ and } v_2 = uv_2'u(uv_1uv_2u)^{p-1}. \] We compute \( \gamma(u) = f, \gamma(v_1) = f_1 \) 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 \( \gamma(v_2) = f_2q_f(f_1q_2f')^{p-1} \) and \( \delta(v_2) = (g_2(v_2)g)^p. \) 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 \((uv_1uv_2u)^p w_2u(uv_1uv_2u)^p \) 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 \( \gamma, \) which yields \((f_1q_1q_2f')^{p-2} = (f_1q_1q_2f')^{p-2}f_2q_f(f_1q_2f)^{2p-1}. \) One may then multiply by \( f_1q_2f \) on the right to get (4), as desired. For \( x, y \in \mathcal{A}^*, \) let \( z_1 = x(uv_1uv_2u)^py \) and \( z_2 = x(uv_1uv_2u)^pw_2u(uv_1uv_2u)^p_\ell. \) 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 \( u' = \varepsilon. \)

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 \( \delta(b) \notin 1_N. \) In particular \( d_2(b) = 0. \) Hence, \( b \) is an infix of length 1 of both \( z_1 \) and \( z_2 \) such that \( d_3(b) < 1. \) 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 \( \delta(z_1) = \delta(z_2) = (x)h \delta(y). \) 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 \( g \notin \beta. \) Again, there are two cases. First, assume that \( \ell = 0. \) Since \( d_2(r) \leq \ell \) by hypothesis, we get \( r \notin 1_N. \) Thus, since \( u \) and \( v_2 \) are infixes of \( z_1 \in \delta^{-1}(r), \) we have \( \delta(u) \notin \beta(v_2) \notin 1_N, \) which exactly says that \( g \notin \beta \notin 1_N. \) 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 \( d_3(u) \geq \ell \) and \( d_3(v_2) \geq \ell, \) \( r \leq_\delta \delta(u) \) and \( r \leq_\delta \delta(v_2). \) In particular, it follows that \( d_3(r) \geq d_3(u) \geq \ell \) and \( d_3(r) \geq d_3(v_2) \geq \ell. \) Since \( d_3(r) \leq \ell \) by hypothesis on \( r, \) we get \( d_3(r) = d_3(u) = d_3(v_2) = \ell. \) Together with \( r \leq_\delta \delta(u) \) and \( r \leq_\delta \delta(v_2), \) this yields \( r \notin \beta(u) \notin \beta(v_2), i.e., r \notin \beta \notin h. \) This completes the proof that \( g \notin \beta. \) Since we also know that \( hg = gh = h, \) we have \( h \leq_\delta g \) and Lemma 1 yields \( g \notin \beta. \) We get \( z \in N \) such that \( g = h. \) 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 \( \delta \)-pairs for \( \beta. \) Moreover, recall that \( u = (u')^p \) where \( p \) is a multiple of \( \omega(M). \) Hence, we have \( \hat{\beta} \) and \( \beta \in E(M). \) Consequently, it follows from Lemma 23 that \( \beta((uv_1uv_2u)^p) = \beta((uv_1uv_2u)^p w_2u(uv_1uv_2u)^p). \) It now suffices to multiply by \( \beta(x) \) on the left and \( \beta(y) \) on the right to obtain \( \hat{\beta}(z_1) = \hat{\beta}(z_2), \) 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 TL[\mathcal{C}_\ell] formula \( \varphi_{s,} \) such that for every \( w \in K_\ell, \) we have \( \hat{\varphi}_{s,} \Leftrightarrow \beta(w) = s. \)

Let us first use Proposition 26 to complete the main proof: we have to show that every language recognized by \( \beta \) belongs to TL[\mathcal{C}_\ell]. Clearly, it suffices to show that \( \beta^{-1}(s) \in TL[\mathcal{C}_\ell] \) for every \( s \in M. \) We apply Proposition 26 for \( \ell = 0. \) Since \( K_0 = B^*, \) this yields a formula \( \varphi_{0, \gamma} \in TL[\mathcal{C}_\ell] \) such that \( L_{\min}(\varphi_{0, \gamma}) = \beta^{-1}(s). \) Thus, \( \beta^{-1}(s) \in TL[\mathcal{C}_0], \) as desired.

It remains to prove Proposition 26. We construct \( \varphi_{s,} \in TL[\mathcal{C}_\ell] \) by induction on \( |N| - \ell. \) If \( \ell = |N|, \) we define \( \varphi_{s,} \) so that \( L_{\min}(\varphi_{s,}) = K_\ell \cap \beta^{-1}(s). \) Since \( K_{|N|} \cap \beta^{-1}(s) \) is finite and TL[\mathcal{C}_\ell] is closed under disjunction, it suffices to build for every word \( w \in B^* \) a TL[\mathcal{C}_\ell] formula \( \varphi_w \) defining \( \{w\}. \) Since \( B^* \in \mathcal{C}_\ell, \) one may use the “F” modality. For \( w = b_1 \cdots b_n, \) let
\[\psi_w = F(b_1 \land F(b_2 \land F(b_3 \land \cdots \land F b_n))).\]

One may then choose \(\varphi_w = \psi_w \land \bigwedge_{u \in B^+ \mid |u| = |w|+1} \neg \psi_u\).

Assume now that \(\ell < |N|\). We present a construction for splitting the words in \(K_\ell\) into two parts: a prefix mapped to an element \(r \in N\) such that \(d_3(r) \leq \ell\) (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 \(i \in \text{Pos}(w) \setminus \{0\}\) 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 \(\pi_{k,u} \in \text{TL}[C_3]\) such that for all \(w \in K_\ell\) and \(i \in \text{Pos}(w) \setminus \{0\}\), \(w, i \models \pi_{k,u} \iff \sigma_k(w, i) = u\).

**Proof.** If \(u = \epsilon\), 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 \(\delta\), so that \(H \models \delta^{-1}(1_N) = C^*\). By definition, \(H \in C_3\). Since \(2 \leq k \leq \ell + 1\), we have \(\ell \geq 1\), which implies, by definition of \(K_\ell\), that no word of \(K_\ell\) can contain a letter \(b\) with \(d_3(b) = 0\). In particular, words of \(K_\ell\) cannot contain letters of \(C\). Therefore, if \(w \in K_\ell\), \(i \in \text{Pos}(w)\) and \(\psi \in \text{TL}[C_3]\), we have \(w, i \models F_H \psi\) 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 \(\pi_{k,u} = (b_1 \land F_H (b_2 \land F_H (b_3 \land \cdots \land F_H b_n)))\).
- If \(n < k\), we let \(\pi_{k,u} = (b_1 \land F_H (b_2 \land F_H (b_3 \land \cdots \land F_H (b_n \land F_H \max)))\).

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 \(i \in \text{Pos}(w)\) is pointed when either \(i \in \{0, |w|+1\}\), or \(i \in \text{Pos}_s(w)\) and \(d_3(\tau_{\ell+1}(w, i)) \geq \ell + 1\).

**Definition 28** (Detection of pointed positions in \(\text{TL}[C_3]\)). Let \(\pi = \min \lor \max \lor \bigvee_{u \in U} \pi_{\ell+1, u}\) where \(U = \{u \in B^+ \mid |u| \leq \ell + 1\} \land d_3(u) \geq \ell + 1\). By definition of \(\pi_{\ell+1, u}\) in Lemma 27, we know that for \(w \in K_\ell\) and \(i \in \text{Pos}(w)\), we have \(w, i \models \pi\) if and only if position \(i\) is pointed.

A position \(i \in \text{Pos}(w)\) which is not pointed is said to be safe. We now prove that we may constrain the evaluation of \(\text{TL}[C_3]\) formulas to infixes that only contain safe positions.

**Lemma 29.** Let \(\psi \in \text{TL}[C_3]\) and \(H \in C_3\). There exist two formulas \(F_H^\psi\) and \(P_H^\psi\) of \(\text{TL}[C_3]\) such that for all \(w \in K_\ell\) and all \(i \in \text{Pos}(w)\), the two following properties hold:

- \(w, i \models F_H^\psi\) if and only if there exists \(j \in \text{Pos}(w)\) such that \(j > i\), \(w, j \models \psi\), \(w(i, j) \in H\) and all positions \(h \in \text{Pos}(w)\) such that \(i < h < j\) are safe.
- \(w, i \models P_H^\psi\) if and only if there exists \(j \in \text{Pos}(w)\) such that \(j < i\), \(w, j \models \psi\), \(w(i, j) \in H\) and all positions \(h \in \text{Pos}(w)\) such that \(j < h < i\) are safe.

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

1. All positions \(h \in \text{Pos}(w)\) such that \(i < h < j\) are safe.
2. Either \(\delta(w(i,j)) = 1_N\) or \(d_3(w(i,j)\tau_{\ell}(w,j)) \leq \ell\).

Assume first that all positions \(h \in \text{Pos}(w)\) such that \(i < h < j\) are safe. If \(i + 1 = j\), then \(w(i,j) = \epsilon\), whence \(\delta(w(i,j)) = 1_N\). Assume now that \(i+1 < j\). Observe that \(w(i,j)\tau_{\ell}(w,j)\) belongs to \(K_\ell\) since it is an infix of \(w \in K_\ell\). Moreover, since \(i+1 < j\), there exists at least one \(h \in \text{Pos}(w)\) 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 \(xyz = w(i,j)\tau_{\ell}(w,j)\) and \(|y| \leq \ell + 1\), we have \(d_3(y) \leq \ell\). Therefore, Lemma 24 entails that \(d_3(w(i,j)\tau_{\ell}(w,j)) \leq \ell\), as desired.
Conversely, assume that either $\delta(w(i,j)) = 1_N$ or $d_3(w(i,j)\sigma(w,j)) \leq \ell$. We start with the latter case. Since $w(i,j)\sigma(w,j) \in K_\ell$, Lemma 24 implies that for every $x,y,z \in B^*$ such that $xyz = w(i,j)\sigma(w,j)$ and $|y| \leq \ell + 1$, we have $d_3(y) \leq \ell$. In particular, it follows that every $h \in \text{Pos}(w)$ such that $i < h < j$ is safe. Assume now that $\delta(w(i,j)) = 1_N$. If $\ell = 0$, then $\sigma(w,j) = \varepsilon$ 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 $\psi \in \text{TL}[C_\delta]$ and $H \in C_\delta$. For every $r \in N$, we let $H_r = H \cap \delta^{-1}(r)$ and $U_r = \{u \in B^* | |u| \leq \ell$ and $d_3(r\delta(u)) \leq \ell\}$. Observe that $H_r \in C_\delta$. Now, in view of the preliminary result, it suffices to define,

$$F^w_H \psi = F_{H_{1_N}} \psi \lor \bigvee_{r \in N} F_{H_r} (\pi_{t,u} \wedge \psi) \quad \text{and} \quad P^w_H \psi = P_{H_{1_N}} \psi \lor \bigvee_{r \in N} (\pi_{t,u} \wedge P_{H_r} \psi).$$

This completes the proof. \hfill △

**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 $b_1, \ldots, b_n \in B$ are those carried by the pointed positions $i_1, \ldots, i_n$. 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 $\tilde{w} \in B^*$ 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 $\beta$ and $\delta$, we know that there is a letter $b_j \in B$ such that $(\beta(b_j-t_j,q_j),\delta(b_j-t_j,q_j)) = (t_j,q_j)$. We let $\tilde{w} = b_{t_1,q_1} \cdots b_{t_n,q_n}$. Note that by definition, $\beta(b_1w_1 \cdots b_nw_n) = \beta(\tilde{w})$ and $\delta(b_1w_1 \cdots b_nw_n) = \delta(\tilde{w})$. Finally, we define a surjective map $i \mapsto \mu(i)$ associating a position $\mu(i) \in \text{Pos}(\tilde{w})$ to each pointed position $i \in \text{Pos}(w)$: 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 $i \in \{0\} \cup \text{Pos}(w)$, one can compute the images of the word $f(w,i)$ under $\beta$ and $\delta$ with a TL$[C_\delta]$ formula. This is where we use Proposition 25.

**Lemma 30.** Let $(t,r) \in M \times N$. There exists $\Gamma_{t,r} \in \text{TL}[C_\delta]$ such that for all $w \in K_\ell$ and all pointed positions $i \in \{0\} \cup \text{Pos}(w)$, we have $w,i \vdash \Gamma_{t,r} \leftrightarrow \beta(f(w,i)) = t$ and $\delta(f(w,i)) = r$.

**Proof.** First observe that by definition, if $w \in K_\ell$ and $i \in \{0\} \cup \text{Pos}(w)$ 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 $d_3(y) \leq \ell$. By Lemma 24, it follows that $d_3(f(w,i)) \leq \ell$. Therefore, if $d_3(r) > \ell$, then $\delta(f(w,i))$ cannot be safe on $r$, and it suffices to define $\Gamma_{t,r} = \bot$.

We now assume that $d_3(r) \leq \ell$. Proposition 25 implies that $K_\ell \cap \beta^{-1}(t) \cap \delta^{-1}(r) \in \text{TL}[C_\delta]$. We get a formula $\psi \in \text{TL}[C_\delta]$ such that for every $w \in K_\ell$, we have $w,0 \vdash \psi$ if and only if $\beta(u) = t$ and $\delta(u) = r$. Using Lemma 29, we modify $\psi$ so that given $w \in K_\ell$, the evaluation of $\psi$ 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 $i \in \{0\} \cup \text{Pos}(w)$ and $j \in \text{Pos}(f(w,i))$, the two following properties hold:

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

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 $\pi \in \text{TL}[C_\delta]$ 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 $\psi = \text{min}$, we let $\langle \psi \rangle_{\min} = \pi$ and $\langle \psi \rangle_{\max} = \bot$. If $\psi = \text{max}$, 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 $\langle \psi' \lor \psi'' \rangle_{\min} = \langle \psi' \rangle_{\min} \lor \langle \psi'' \rangle_{\min}$, $\langle \psi' \land \psi'' \rangle_{\min} = \langle \psi' \rangle_{\min} \land \langle \psi'' \rangle_{\min}$ and $\langle \neg \psi' \rangle_{\min} = \neg \langle \psi' \rangle_{\min}$, and similarly for $\langle \cdot \rangle_{\max}$.
We have to show that this is where we use induction in Proposition 26. Indeed, we have the following lemma.

\[ \psi \in F_H \psi' \text{ for } H \in C_\delta, \text{ let } \langle \psi \rangle_{\min} = F_H^\alpha \langle \psi \rangle_{\max} \text{ and } \langle \psi \rangle_{\max} = \neg \pi \land F_H^\alpha \langle \psi \rangle_{\max}. \]

Symmetrically, if \( \psi = F_H \psi' \) for some \( H \in C_\delta \), we define \( \langle \psi \rangle_{\min} = \neg \pi \land P_H^\alpha \langle \psi \rangle_{\min} \) and \( \langle \psi \rangle_{\max} = P_H^\alpha \langle \psi \rangle_{\min}. \) 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 \( \zeta_s \in TL[C_\delta] \) such that for every \( w \in K_\ell \), we have \( w, 0 \models \zeta_s \iff \beta(\tilde{w}) = s \). Given \( s \in M \), it will then suffice to define \( \varphi_{\ell,s} \in TL[C_\delta] \) as follows:

\[
\varphi_{\ell,s} = \bigvee_{\{(s_1,s_2) \in M^2|s_1=s_2\}} \left( \bigvee_{r \in N} \Gamma_{s_1,r} \land \zeta_{s_2} \right).
\]

Indeed, it is straightforward that for every word \( w \in K_\ell \), we have \( \beta(w) = \beta(f(w,0)) \beta(\tilde{w}) \). Consequently, by definition of \( \varphi_{\ell,s} \), we get \( w, 0 \models \varphi_{\ell,s} \iff \beta(f(w,0)) \beta(\tilde{w}) = s \iff \beta(w) = s \) for all \( w \in K_\ell \), which concludes the proof of Proposition 26. We now concentrate on building \( \zeta_s \).

This is where we use induction in Proposition 26. Indeed, we have the following lemma.

\begin{lemma}
For every \( w \in K_\ell \), we have \( \tilde{w} \in K_{\ell+1} \).
\end{lemma}

\begin{proof}
Let \( k \leq \ell + 1 \) and \( x,y,z \in B^* \) such that \( \tilde{w} = xyz \) and \( |y| = k \). We have to prove that \( d_3(y) \geq k \). Let \( w = w_0b_1w_1 \cdots b_nw_n \) be the pointed decomposition of \( w \). By definition of \( \tilde{w} \), we have \( \delta(y) = \delta(b_1w_1 \cdots b_{h+1}w_{h+1} - 1) \) for some \( h \leq n \). Let \( u = b_hw_h \cdots b_{h+k-1}w_{h+k-1} \).

We have to show that \( d_3(y) = d_3(u) \geq k \). Clearly, \( |u| \geq k \). Hence, if \( k \leq \ell \), the hypothesis that \( w \in K_\ell \) yields \( d_3(u) \geq k \). Otherwise, \( k = \ell + 1 \). Thus, \( |u| \geq \ell + 1 \) and since the position labeled by \( b_h \) in \( w \) is pointed, this yields \( d_3(u) \geq \ell + 1 \). In both cases, we get \( d_3(y) \geq k \).
\end{proof}

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

We construct \( \psi \) by structural induction on \( \psi \). If \( \psi \in \{min, max, T, \top, \bot\} \), we let \( \psi_s = \psi \). Suppose now that \( \psi = b_{k,q} \in B \) for \( (t,q) \in P \). Then, when evaluated in \( w \) at a pointed position \( i \) carrying a \( "b" \), we want \( \psi \) to check that \( \beta(b) \beta(f(w, i)) = t \) and \( \delta(b) \beta(f(w, i)) = q \). Let \( T = \{ (b,t',q') \in B \times M \times N | \beta(b) = t \text{ and } \delta(b) = q \} \). Using the formulas \( \Gamma_t \) from Lemma 30, we define \( \psi = \bigvee_{(b',t',q') \in T} (b \land \Gamma_t \land \Gamma_{t'} \land \Gamma_{t'^{2n}}) \). Boolean operators are handled as expected.

It remains to deal with temporal modalities, i.e., the case where there exists \( H \in C_\delta \) such that \( \psi = F_H \psi' \) or \( \psi = P_H \psi' \). For every \( b \in B \), let \( F_b = \{ t \in N | \delta(b) \in \delta(H) \} \). We define:

\[
[F_H \psi']_{\psi'} = \begin{cases} 
F_{B}^{n} \pi \land (\bigvee_{b \in B} (b \land F_{b}^{n-1} (\pi \land \psi') )) & \text{if } \epsilon \notin H, \\
F_{B}^{n} \pi \land (\bigvee_{b \in B} (b \land F_{b}^{n-1} \psi' )) & \text{if } \epsilon \in H.
\end{cases}
\]

\[
[P_H \psi']_{\psi'} = \begin{cases} 
F_{B}^{n} \pi \land (\bigvee_{b \in B} (b \land F_{b}^{n-1} (\pi \land \psi') )) & \text{if } \epsilon \notin H, \\
F_{B}^{n} \pi \land (\bigvee_{b \in B} (b \land F_{b}^{n-1} \psi' )) & \text{if } \epsilon \in H.
\end{cases}
\]

We give an intuition when \( \psi = F_H \psi' \) and \( \epsilon \notin H \). Let \( w \in B^* \) be the pointed decomposition of \( w \) and \( \tilde{w} = b_k' \cdots b_1' \). Let \( i_k \in Pos(w) \) be the position of the distinguished \( b_k \), so that \( \mu(i_k) = k \). Now, \( \tilde{w}, k \models F_H \psi' \) when there exists \( m > k \) such that \( \tilde{w}, m \models \psi' \) and \( b_{k+1} \cdots b_{m-1} \in H \). The construction ensures that \( w, i_k \models [F_H \psi'] \) when there exists \( m > k \) such that \( w, m \models \psi' \) and \( b_{k+1} \cdots b_{m-1} \in H \). The purpose of using \( F_{b_{k+1}}(\pi \land \ldots) \) 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 \( w, i \models [\psi] \Leftrightarrow \tilde{w}, \mu(i) \models \psi \) for all \( w \in K_\ell \) and all pointed positions \( i \in Pos(w) \). This concludes the proof.
6 Natural restrictions of generalized unary temporal logic

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

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

\textbf{Remark 32.} Note that FL[\mathcal{C}] formulas are evaluated at the leftmost unlabeled position whereas 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 FL(\mathcal{C}) and 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 deterministic right deterministic marked products

We now establish a connection between the operators \mathcal{C} \mapsto FL(\mathcal{C}) and \mathcal{C} \mapsto PL(\mathcal{C}) and the two weaker variants RPol and LPol of unambiguous polynomial closure. Consider a marked product \L_{a_1}L_1 \cdots a_n L_n. For 1 \leq i \leq n, we write \H_i = L_1 a_1 L_2 \cdots a_{i-1} L_{i-1} and \K_i = L_i a_{i+1} L_{i+1} \cdots a_n L_n. We say that \L_{a_1}L_1 \cdots a_n L_n is right deterministic (resp. left deterministic) when we have \A^* a_i \K_i \cap \H_i = \emptyset (resp. \H_i a_i \A^* \cap \H_i = \emptyset) for every 1 \leq i \leq n. The right polynomial closure of a class \mathcal{C}, written RPol(\mathcal{C}), consists of all finite disjoint unions of right deterministic marked products \L_{a_1}L_1 \cdots a_n L_n such that \L_0, \ldots, L_n \in \mathcal{C} (by “disjoint” we mean that the languages in the union must be pairwise disjoint). Similarly, the left polynomial closure LPol(\mathcal{C}) of \mathcal{C} consists of all finite disjoint unions of left deterministic marked products \L_{a_1}L_1 \cdots a_n L_n such that \L_0, \ldots, 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 RPol(\mathcal{C}) and LPol(\mathcal{C}).

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

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

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

\textbf{Proposition 34.} For every prevariety \mathcal{C}, we have LPol(BPol(\mathcal{C})) \subseteq PL(\mathcal{C}).
Then, Proposition 35.

Let $C$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \rightarrow M$ be its syntactic morphism. Assume that $\alpha$ satisfies the following property:

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

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

Proposition 36.

Let $C$ be a prevariety, $L \subseteq A^*$ be a regular language and $\alpha : A^* \rightarrow M$ be its syntactic morphism. Assume that $\alpha$ satisfies the following property:

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

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

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 $\text{RPol}(D)$ proved in [21]. Let us first present it. For every class $D$, we define a preorder $\preceq_D$ and an equivalence $\sim_D$ over $M$. Given $s, t \in M$, we let,

$$
s \sim_D t \quad \text{if and only if} \quad s \in F \iff t \in F \text{ for every } F \subseteq M \text{ such that } \alpha^{-1}(F) \in D,
$$

$$
s \preceq_D t \quad \text{if and only if} \quad s \in F \Rightarrow t \in F \text{ for every } F \subseteq M \text{ such that } \alpha^{-1}(F) \in D.
$$

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

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

Moreover, it is shown in [27, Lemma 6.6] that $\preceq_{\text{co-Pol}(C)}$ 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 $C$-pair, then $xey \preceq_{\text{co-Pol}(C)} xey$ (the proof is based on the algebraic characterization of Pol($C$), see [23]). This yields $s_0, \ldots, s_n \in M$ 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 $C$-pair, $s_{i-1} = xeyq$ and $s_i = xey$. We use induction on $i$ to prove that $s_i^{\omega+1} = s_is^{\omega}$ for every $i \leq n$. Since $s_n = t$, the case $i = n$ yields the desired result. When $i = 0$, it is immediate that $s_0^{\omega+1} = s_0s^{\omega}$ since $s_0 = s$. Assume now that $i \geq 1$. 

\text{CSL 2024}
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, \( s_{i-1} = xeqy \) and \( s_i = xey \). Since \((s^{\omega + 1})^{\omega + 2} = xeqys^{\omega + 2}\), we get \( s^{\omega + 2} = (xeqys^{\omega + 2})^{\omega + 2} \). Hence, we get

\[
\begin{align*}
\omega \quad s^{\omega + 2} &= x(eqys^{\omega + 2}q)eys^{\omega + 1} \\
\omega &= xeqys^{\omega + 2}qxeys^{\omega} \quad \text{by (5) since \((e, q)\) is a \( \mathcal{C}\)-pair} \\
\omega &= xeys^{\omega}xeqys^{\omega.}
\end{align*}
\]

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

### 6.2 Statements

The classes \( \mathcal{F}_L(\mathcal{C}) \) and \( \mathcal{P}_L(\mathcal{C}) \) admit algebraic characterizations similar to that of \( \mathcal{F}_L(\mathcal{C}) \). 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 \( \mathcal{F}_L(\mathcal{C}) \) and \( \mathcal{P}_L(\mathcal{C}) \).

#### Theorem 37

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

#### Theorem 38

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

Since \( \mathcal{F}_L(\mathcal{C}) \) and \( \mathcal{P}_L(\mathcal{C}) \) are symmetrical, it is natural to consider a third class denoted \( \mathcal{F}_L(\mathcal{C}) \cap \mathcal{P}_L(\mathcal{C}) \). It consists of all languages belonging simultaneously to \( \mathcal{F}_L(\mathcal{C}) \) and \( \mathcal{P}_L(\mathcal{C}) \). 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 \( \alpha : A^* \rightarrow M \) be its syntactic morphism. The two following properties are equivalent:
1. \( L \in \mathcal{F}_L(\mathcal{C}) \cap \mathcal{P}_L(\mathcal{C}) \).
2. Every \( \mathcal{C}\)-orbit for \( \alpha \) is \( \mathcal{J}\)-trivial.

Recall that given a regular language \( L \subseteq A^* \) as input, its syntactic morphism \( \alpha : A^* \rightarrow M \) can be computed. Moreover, Lemma 5 implies that all \( \mathcal{C}\)-orbits for \( \alpha \) 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 \( \mathcal{F}_L(\mathcal{C}), \mathcal{P}_L(\mathcal{C}) \) and \( \mathcal{F}_L(\mathcal{C}) \cap \mathcal{P}_L(\mathcal{C}) \) 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 $\text{TL}(\mathcal{C})$, $\text{FL}(\mathcal{C})$ and $\text{PL}(\mathcal{C})$. 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 $\text{TL}(\mathcal{C})$, $\text{FL}(\mathcal{C})$ and $\text{PL}(\mathcal{C})$ 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 $\text{TL}_0(\mathcal{C}) = \mathcal{C}$ and level $n \geq 1$ is $\text{TL}_n(\mathcal{C}) = \text{TL}(\text{TL}_{n-1}(\mathcal{C}))$. It can be shown that the hierarchies of bases $\text{ST} = \{\emptyset, A^+\}$ and $\text{DD} = \{\emptyset, \{\varepsilon\}, A^+, A^*\}$ are strict. Thus, since $\text{BPo}(\mathcal{C}) \subseteq \text{TL}(\mathcal{C})$, 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 $\text{TL}_2(\text{ST})$ and $\text{TL}_2(\text{DD})$ have decidable separation, then $\text{TL}_3(\text{ST})$ and $\text{TL}_3(\text{DD})$ 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 $\mathcal{C} \mapsto \text{TL}(\mathcal{C})$, $\text{FL}(\mathcal{C})$ and $\text{PL}(\mathcal{C})$.

References


