Separation and the Successor Relation*

Thomas Place and Marc Zeitoun

LaBRI, Bordeaux University, France, `firstname.lastname@labri.fr`.

Abstract

We investigate two problems for a class $\mathcal{C}$ of regular word languages. The $\mathcal{C}$-membership problem asks for an algorithm to decide whether an input language belongs to $\mathcal{C}$. The $\mathcal{C}$-separation problem asks for an algorithm that, given as input two regular languages, decides whether there exists a third language in $\mathcal{C}$ containing the first language, while being disjoint from the second. These problems are considered as means to obtain a deep understanding of the class $\mathcal{C}$.

It is usual for such classes to be defined by logical formalisms. Logics are often built on top of each other, by adding new predicates. A natural construction is to enrich a logic with the successor relation. In this paper, we obtain simple self-contained proofs of two transfer results: we show that for suitable logically defined classes, the membership, resp. the separation problem for a class enriched with the successor relation reduces to the same problem for the original class.

Our reductions work both for languages of finite words and infinite words. The proofs are mostly self-contained, and only require a basic background on regular languages. This paper therefore gives new, simple proofs of results that were considered as difficult, such as the decidability of the membership problem for the levels 1, 3/2, 2 and 5/2 of the dot-depth hierarchy.

1998 ACM Subject Classification F.4.3 Formal Languages

Keywords and phrases Separation Problem, Regular Word Languages, Logics, Decidable Characterizations, Semidirect Product

1 Introduction

Context. A central problem in the theory of formal languages is to characterize and understand the expressive power of high level specification formalisms. Monadic second order logic (MSO) is such a formalism, which is both expressive and robust. For several classes of structures, such as words or trees, it has the same expressive power as finite automata and defines the class of regular languages. In this paper, we investigate fragments of MSO over words. In this context, understanding the expressive power of a fragment is associated to two decision problems: the *membership problem* and the *separation problem*.

For a fixed logical fragment $\mathcal{F}$, the *$\mathcal{F}$-membership problem* asks for a decision procedure that tests whether some input regular language can be expressed by a formula from $\mathcal{F}$. To obtain such an algorithm, one has to consider and understand *all* properties that can be expressed within $\mathcal{F}$, which requires a deep understanding of the fragment $\mathcal{F}$. On the other hand, the *$\mathcal{F}$-separation problem* is more general. It asks for a decision procedure that tests whether given *two* input regular languages, there exists a third one in $\mathcal{F}$ containing the first language while being disjoint from the second one. Since regular languages are closed under complement, membership reduces to separation: a language is in $\mathcal{F}$ iff it can be separated from its complement. Usually, the separation problem is more difficult than the membership problem but also more rewarding wrt. the knowledge gained on the investigated fragment $\mathcal{F}$.

These two problems have been considered and solved for many natural fragments of MSO. Among these, the most prominent one is first-order logic FO(<) equipped with a

* Supported by ANR 2010 BLAN 0202 01 FREC

predicate $<$ for the linear ordering. The solution to the membership problem, known as the McNaughton-Papert-Schützenberger Theorem [22, 12], has been revisited until recently [6]. The theorem states that a regular language is definable in FO($<$) if and only if its *syntactic semigroup* is aperiodic. The syntactic semigroup is a finite algebraic object that can be computed from any regular language. Since aperiodicity can be defined as an equation that needs to be satisfied by all of its elements, this yields decidability of FO($<$)-definability. This result now serves as a template, which is commonly followed in this line of research.

The separation problem has also been successfully solved for first-order logic [8]. Actually, the problem was first addressed in a purely algebraic framework, and was later identified as equivalent to our separation problem [2]. As for membership, this problem is still revisited today and a new self-contained and combinatorial proof was obtained in [19].

Motivation. We are interested in natural fragments of FO($<$) obtained by restricting either the number of variables or the number of quantifier alternations allowed in formulas. Such restrictions in general give rise to several variants of the same fragment. Indeed, in most cases, the drop in expressive power forbids the use of natural relations that could be defined from the linear order in FO($<$). The main example considered in this paper is $+1$, the *successor relation*, together with predicates *min* and *max* for the first and last positions in a word. This means that one can define two distinct variants of the same fragment depending on whether we decide to explicitly add these predicates in the signature or not. An example is the fragment Σ_n, which consists of first-order formulas whose prenex normal form has at most $(n-1)$ quantifier alternations and starts with an existential block. Since defining $+1$ requires an additional quantifier alternation, $\Sigma_n(<, +1, min, max)$ has indeed stronger expressiveness than $\Sigma_n(<)$. The motivation of this paper is to obtain decidability results for such enriched fragments.

State of the Art. Even when the weak fragment is known to have decidable membership, proving that the enriched one has the same property can be nontrivial. Examples include the membership proofs of $\mathcal{B}\Sigma_1(<, +1, min, max)$ (Boolean combinations of $\Sigma_1(<, +1, min, max)$ formulas) and $\Sigma_2(<, +1)$, which require difficult and intricate combinatorial arguments [9, 7, 10] or a wealth of algebraic machinery [14, 16]. Another issue is that most proofs directly deal with the enriched fragment. Given the jungle of such logical fragments, it is desirable to avoid such an approach, treating each variant of the same fragment independently. Instead, a satisfying approach is to first obtain a solution of the membership and separation problems for the less expressive variant and then to lift it to other variants via a generic transfer result.

This approach has first been investigated by Straubing for the membership problem [26] in an algebraic framework, and later adapted to be able to treat classes not closed under complement [16]. Transferring the logical problem to this algebraic framework requires preliminary steps, still *specific* to the investigated class, to prove that: **(1)** a language is definable in the fragment iff its syntactic semigroup belongs to a specific algebraic variety V (*e.g.*, the variety of aperiodic monoids for FO($<$)), and **(2)** membership to V is decidable. Next, though this is not immediate, for most fragments of FO($<$), it has been proved that **(3)** when the weaker variant corresponds to a variety V, the variant with successor corresponds to the variety $\mathsf{V} * \mathsf{D}$, built generically from V. Hence, Straubing's approach was to prove that **(4)** the operator $\mathsf{V} \mapsto \mathsf{V} * \mathsf{D}$ preserves decidability. Unfortunately, this is not true in general [3]. Actually, while decidability is preserved for all known logical fragments, there is no generic result that captures them all. In particular, for the less expressive fragments, one has to use completely *ad hoc* proofs. In the separation setting, things behave well: it has been shown that decidability of separation is preserved by the operation $\mathsf{V} \mapsto \mathsf{V} * \mathsf{D}$ [24]. While interesting when already starting from algebra, this approach has several downsides:

- Dealing with algebra hides the logical intuitions, while our primary goal is to understand the expressiveness of logics.
- Going from logic to algebra requires to be acquainted with new notions and vocabulary, as well as involved theoretical tools. Proofs are also often nontrivial and require a deep understanding of complex objects, which may be scattered in the bibliography.
- Despite step **(4)** which is generic to some extent, arguments specific to the investigated class are pushed to steps **(1)**–**(3)**, and they are often nontrivial.

Contributions. We give a new proof that decidability of separation can be transferred from a weak to an enriched fragment. We present the result in two different forms.

The first one is non-algebraic: we work directly with the logical fragments, without using varieties. The transfer result is generic and its proof mostly is: the only specific argument is an Ehrenfeucht-Fraïssé game that can be adapted to all natural fragments with minimal difficulty (we prove it in the long version for all considered fragments). The benefits are that: **(1)** this new proof is self-contained and much simpler than previous ones. It only relies on two basic, well-known notions: recognizability by semigroups and Ehrenfeucht-Fraïssé games; **(2)** it works with classes which are not closed under complement, contrary to [24]. This allows us to capture the Σ and Π levels in the quantifier alternation hierarchy of first-order logic; **(3)** under an additional hypothesis on the logical fragment, which is met for most fragments we investigate, the decidability result of the separation problem also extends to the membership problem; and **(4)** the proof adapts smoothly to infinite words using the notion of ω-semigroups, as shown in the long version of this paper.

The second form is algebraic and generic: we work with varieties and prove that $\mathsf{V} \mapsto \mathsf{V} * \mathsf{D}$ preserves the decidability of separation, hence giving an elementary proof of a result of [24]. Even in this algebraic form, we completely bypass involved constructions or notions, such as pointlike sets for categories developed in [24], thus making the proof accessible.

As corollaries, since $\mathcal{B}\Sigma_1(<)$ and $\Sigma_2(<)$ both enjoy decidable separation [5, 21, 18], we obtain that this is also the case for the fragments $\mathcal{B}\Sigma_1(<, +1, min, max)$ and $\Sigma_2(<, +1)$, known as levels 1 and 3/2 of the dot-depth hierarchy. These new results strengthen the previous ones [9, 7] that showed decidability of membership and were considered as difficult. We actually obtain that separation for $\Sigma_n(<, +1, min, max)$ reduces to separation for $\Sigma_n(<)$. Since we also transfer decidability of the membership problem, and since the fragments $\mathcal{B}\Sigma_2(<)$ of Boolean combinations of $\Sigma_2(<)$ formulas and $\Sigma_3(<)$ have decidable membership [18] we deduce that the same holds for $\mathcal{B}\Sigma_2(<, +1)$ and $\Sigma_3(<, +1)$, known as levels 2 and 5/2 of the dot-depth hierarchy.

Organization of the Paper. In Section 2, we set up the notation and we present the separation problem and the main logics we deal with. Section 3 is devoted to our main tool: languages of well-formed words. In Section 4, we use it to prove our transfer results for all fragments from the logical perspective, and in Section 5, we show that decidability of the separation problem for the variety V entails the same for $\mathsf{V} * \mathsf{D}$.

2 Preliminaries

In this section we provide preliminary definitions on regular languages and separation.

Words, Languages. We fix a finite alphabet A. Let A^+ be the set of all nonempty finite words and A^* be the set of all finite words over A. If u, v are words, we denote by $u \cdot v$ or by uv the word obtained by concatenating u and v. For convenience, we only consider wlog. languages that do not contain the empty word. That is, a language is a subset of A^+. We work with regular languages, *i.e.*, languages definable by *nondeterministic finite automata* (NFA).

Separation. Given three languages K, L, L', we say that K *separates* L from L' if

$$L \subseteq K \text{ and } K \cap L' = \emptyset.$$

If $\mathcal{C}$ is a class of languages, we say that L is $\mathcal{C}$*-separable* from L' if there exists $K \in \mathcal{C}$ that separates L from L'. Note that if $\mathcal{C}$ is closed under complement, L is $\mathcal{C}$-separable from L' iff L' is $\mathcal{C}$-separable from L. However, this is not true for a class $\mathcal{C}$ not closed under complement, such as the classes $\Sigma_n(<)$ of the quantifier alternation hierarchy, which we shall consider.

Given a class $\mathcal{C}$, the $\mathcal{C}$-separation problem asks for an algorithm which, given as input two regular languages L, L', decides whether L is $\mathcal{C}$*-separable* from L'. The $\mathcal{C}$-membership problem, which asks whether an input regular language belongs to $\mathcal{C}$, reduces to the $\mathcal{C}$-separation problem, as a regular language belongs to $\mathcal{C}$ iff it is $\mathcal{C}$-separable from its complement.

Logics. We investigate several fragments of first-order logic on finite words. We view a finite word as a logical structure made of a sequence of positions labeled over A. We work with first-order logic FO($<$) using a unary predicate P_a for each $a \in A$, which selects positions labeled with an a, as well as binary predicates '$=$' for equality and '$<$' for the linear order. Such a formula defines the regular language of all words that satisfy it. We will freely use the name of a logical fragment of FO($<$) to denote the class of languages definable in this fragment. Observe that FO($<$) is powerful enough to express the following logical relations:

- First position, $min(x)$: $\quad \forall y\ \neg(y < x)$.
- Last position, $max(x)$: $\quad \forall y\ \neg(x < y)$.
- Successor, $y = x + 1$: $\quad x < y \wedge \neg(\exists z\ x < z \wedge z < y)$.

However, for most fragments of FO($<$) this is not the case. For example, in the two-variables restriction $\mathrm{FO}^2(<)$ of FO($<$), it is not possible to express successor, as it requires quantifying over a third variable. For these fragments $\mathcal{F}$, adding the predicates min, max and $+1$ yields a strictly more powerful logic $\mathcal{F}^+$. Our goal is to prove a transfer result for such fragments: given a fragment, if the separation problem is decidable for the weak variant $\mathcal{F}$, then it is decidable as well for the strong variant $\mathcal{F}^+$ obtained by enriching $\mathcal{F}$ with the above relations. The technique is *generic*, meaning that it is not bound to a particular logic. In particular, our transfer result applies to the following well-known logical fragments:

- FO($=$), the restriction of FO($<$) in which the linear order cannot be used, and only equality between two positions can be tested. The enriched fragment FO($=, +1$) (min and max can be defined in the logic) defines locally threshold testable languages [30].
- All levels in the quantifier alternation hierarchy of first-order logic. A first-order formula is $\Sigma_i(<)$ (resp. $\Pi_i(<)$) if its prenex normal form contains at most $(i-1)$ quantifier alternations and starts with an $\exists$ (resp. a $\forall$) quantifier block. Finally, a $\mathcal{B}\Sigma_i(<)$ formula is a boolean combination of $\Sigma_i(<)$ and $\Pi_i(<)$ formulas.
  Since for all fragments above $\Sigma_2(<)$, a formula involving min and max can be expressed without these predicates in the same logic, we shall denote the enriched fragments by $\Sigma_1(<, +1, min, max)$, $\mathcal{B}\Sigma_1(<, +1, min, max)$, and then by $\Sigma_2(<, +1)$, $\mathcal{B}\Sigma_2(<, +1)$, ...
- $\mathrm{FO}^2(<)$, the restriction of FO($<$) using only two reusable variables. The corresponding enriched fragment is $\mathrm{FO}^2(<, +1)$, since min and max can again be expressed in the logic.

Figure 1 summarizes all fragments the technique applies to. We prove the following theorem:

▶ **Theorem 1.** *Let $\mathcal{F}$ and $\mathcal{F}^+$ be respectively the weak and strong variants of one of the logical fragments in Figure 1. Then $\mathcal{F}^+$-separability can be effectively reduced to $\mathcal{F}$-separability.*

Weak variant	FO(=)	$\mathrm{FO}^2(<)$	$\Sigma_i(<)$	$\mathcal{B}\Sigma_i(<)$
Strong variant	FO(=, +1)	$\mathrm{FO}^2(<,+1)$	$\Sigma_i(<,+1,min,max)$	$\mathcal{B}\Sigma_i(<,+1,min,max)$

Figure 1 Logical fragments to which the technique applies

All these logical fragments have a rich history and have been extensively studied in the literature. In particular, the separation problem is known to be decidable for the following fragments: FO(=), $\mathrm{FO}^2(<)$, $\Sigma_1(<)$, $\mathcal{B}\Sigma_1(<)$, $\Sigma_2(<)$ [5, 21, 18]. This means that, from our results, we obtain decidability of separation for FO(=, +1), $\mathrm{FO}^2(<,+1)$, $\Sigma_1(<,+1,min,max)$, $\mathcal{B}\Sigma_1(<,+1,min,max)$ and $\Sigma_2(<,+1)$. Note that for FO(=, +1), $\mathrm{FO}^2(<,+1)$ and $\mathcal{B}\Sigma_1(<,+1,min,max)$, the results could already be obtained as corollaries of algebraic theorems of Steinberg [24] and Almeida [2]. As explained in the introduction, an issue with this approach is that the proof of Steinberg's result relies on deep algebraic arguments and is not tailored to separation (the connection with separation is made by Almeida [2]). For $\Sigma_1(<,+1,min,max)$ and $\Sigma_2(<,+1)$, the result is new, as Steinberg's result does not apply to classes of languages that are not closed under complement.

3 Tools: Semigroups and Morphism of Well-Formed Words

In this section, we define the main tools used in the paper. First, we recall the well-known semigroup based definition of regular languages: a language is regular if and only if it can be recognized by a finite semigroup. Our second tool, *well-formed words*, is specific to our problem and plays a key role in our transfer result.

3.1 Semigroups and Monoids

We work with the algebraic representation of regular languages in terms of semigroups. A *semigroup* is a set S equipped with an associative product, written $s \cdot t$ or st. A *monoid* is a semigroup S having a neutral element 1_S, *i.e.*, such that $s \cdot 1_S = 1_S \cdot s = s$ for all $s \in S$. If S is a semigroup, then S^1 denotes the monoid $S \cup \{1_S\}$ where $1_S \notin S$ is a new element, acting as neutral element. Note that we add such a new identity even if S is already a monoid.

An element $e \in S$ is *idempotent* if $e \cdot e = e$. We denote by $E(S)$ the set of idempotents of S. Given a *finite* semigroup S, it is folklore and easy to see that there is an integer $\omega(S)$ (denoted by ω when S is understood) such that for all s of S, s^ω is idempotent: $s^\omega = s^\omega s^\omega$.

Note that A^+ and A^* equipped with concatenation are respectively a semigroup and a monoid called the *free semigroup over* A and the *free monoid over* A. Let $L \subseteq A^+$ be a language and S be a semigroup (resp. monoid). We say that L is *recognized by* S if there exist a morphism $\alpha : A^+ \to S$ (resp. $\alpha : A^* \to S$) and a set $F \subseteq S$ such that $L = \alpha^{-1}(F)$.

Semigroups and Separation. The separation problem takes as input two regular languages L, L'. It is convenient to work with a single object recognizing both of them, rather than having to deal with two. Let S, S' be semigroups recognizing L, L' together with the associated morphisms α, α', respectively. Clearly, L and L' are both recognized by $S \times S'$ with the morphism $\alpha \times \alpha' : A^+ \to S \times S'$ mapping w to $(\alpha(w), \alpha'(w))$. From now on, we work with such a single semigroup recognizing both languages. Replacing $S \times S'$ with its image under $\alpha \times \alpha'$, one can also assume that this morphism is surjective. To sum up, we assume from now on, wlog., that L and L' are recognized by a single surjective morphism.

3.2 Well-Formed Words

In this section, we define our main tool for this paper. Assume that $\mathcal{F}$ is the weak variant of one of the logical fragments of Figure 1 and let $\mathcal{F}^+$ be the corresponding strong variant. To any semigroup morphism $\alpha : A^+ \to S$ into a finite semigroup S, we associate a new alphabet $\mathbb{A}_\alpha$ called the alphabet of *well-formed words*. The main intuition behind this notion is that the $\mathcal{F}^+$-separation problem for any two regular languages recognized by α can be reduced to the $\mathcal{F}$-separation problem for two regular languages over $\mathbb{A}_\alpha$.

The alphabet $\mathbb{A}_\alpha$, called *alphabet of well-formed words of* α, is defined from $\alpha : A^+ \to S$ by:

$$\mathbb{A}_\alpha = (E(S) \times S \times E(S)) \quad \cup \quad (S \times E(S)) \quad \cup \quad (E(S) \times S) \quad \cup \quad S.$$

We will not be interested in all words of $\mathbb{A}_\alpha^+$, but only in those that are well-formed. A word $\mathbb{w} \in \mathbb{A}_\alpha^+$ is said to be *well-formed* if one of the following two properties holds:

- $\mathbb{w}$ is a single letter $s \in S$,
- $\mathbb{w}$ has length $\geqslant 2$ and is of the form

$$(s_0, f_0)\cdot(e_1, s_1, f_1)\cdots(e_n, s_n, f_n)\cdot(e_{n+1}, s_{n+1}) \in (S\times E(S))\cdot(E(S)\times S\times E(S))^*\cdot(E(S)\times S)$$

with $f_i = e_{i+1}$ for $0 \leqslant i \leqslant n$.

▶ **Fact 2.** *The set of well-formed words of $\mathbb{A}_\alpha^+$ is a regular language.*

We now define a morphism $\beta : \mathbb{A}_\alpha^+ \to S$ as follows. If $s \in S$, we set $\beta(s) = s$, if $(e, s) \in E(S) \times S$, we set $\beta((e, s)) = es$, if $(s, e) \in S \times E(S)$, we set $\beta((s, e)) = se$ and if $(e, s, f) \in E(S) \times S \times E(S)$, we set $\beta((e, s, f)) = esf$.

Associated Language of Well-formed Words. To any language $L \subseteq A^+$ that is recognized by α, one associates a language of well-formed words $\mathbb{L} \subseteq \mathbb{A}_\alpha^+$:

$$\mathbb{L} = \big\{\mathbb{w} \in \mathbb{A}_\alpha^+ \mid \mathbb{w} \text{ is well-formed and } \beta(\mathbb{w}) \in \alpha(L)\big\}.$$

By definition, the language $\mathbb{L} \subseteq \mathbb{A}_\alpha^+$ is the intersection of the language of well-formed words with $\beta^{-1}(\alpha(L))$. Therefore, it is immediate by Fact 2 that it is regular, more precisely:

▶ **Fact 3.** *Let $L \subseteq A^+$ that is recognized by α. Then the associated language of well-formed words $\mathbb{L} \subseteq \mathbb{A}_\alpha^+$ is a regular language that one can effectively compute from a recognizer of L.*

4 Logical Approach

In this section, we prove Theorem 1 from a logical perspective. We begin with presenting our 'separation' theorem, which will entail the 'membership' theorem as a simple consequence.

▶ **Theorem 4.** *Let $\mathcal{F}$ and $\mathcal{F}^+$ be respectively the weak and strong variants of one of the logical fragments in Figure 1.*

Let L, L' be two languages recognized by a morphism $\alpha : A^+ \to S$ into a finite semigroup S. Let $\mathbb{L}, \mathbb{L}' \subseteq \mathbb{A}_\alpha^+$ be the languages of well-formed words associated with L, L', respectively. Then L is $\mathcal{F}^+$-separable from L' iff $\mathbb{L}$ is $\mathcal{F}$-separable from $\mathbb{L}'$.

Theorem 4 reduces $\mathcal{F}^+$-separation to $\mathcal{F}$-separation. The latter was already known to be decidable for several weak variants in Figure 1, namely for FO(=) [17], $\mathrm{FO}^2(<)$ [21], $\Sigma_1(<)$ [5], $\mathcal{B}\Sigma_1(<)$ [5, 21] and $\Sigma_2(<)$ [18]. Hence, we get the following corollary.

▶ **Corollary 5.** *Let L, L' be regular languages. Then the following problems are decidable:*

- *whether L is $\mathrm{FO}(=,+1)$-separable from L'.*
- *whether L is $\mathrm{FO}^2(<,+1)$-separable from L'.*
- *whether L is $\Sigma_1(<,+1,min,max)$-separable from L'.*
- *whether L is $\mathcal{B}\Sigma_1(<,+1,min,max)$-separable from L'.*
- *whether L is $\Sigma_2(<,+1)$-separable from L'.*

Notice that since the membership problem reduces to the separation problem, this also gives a new proof that all these fragments have a decidable membership problem. This is of particular interest for $\mathrm{FO}^2(<,+1)$, $\mathcal{B}\Sigma_1(<,+1,min,max)$ and $\Sigma_2(<,+1)$ for which the previous proofs [26, 1, 20], [9], [7, 16, 14] are known to be quite involved. It turns out that for $\Sigma_2(<,+1)$, we can do even better and entirely avoid separation. Indeed, when $\mathcal{F}$ is expressive enough, Theorem 4 can be used to prove a similar theorem for the membership problem.

▶ **Theorem 6.** *Let $\mathcal{F}$ and $\mathcal{F}^+$ be respectively the weak and strong variants of one of the logical fragments in Figure 1. Moreover, assume that for any alphabet of well-formed words, the set of well-formed words over this alphabet is definable in $\mathcal{F}$.*

Let L be a language recognized by a morphism $\alpha : A^+ \to S$ into a finite semigroup S. Let $\mathbb{L} \subseteq \mathbb{A}_\alpha^+$ be the language of well-formed words associated with L. Then L is definable in $\mathcal{F}^+$ iff $\mathbb{L}$ is definable in $\mathcal{F}$.

Proof. Set $K = A^+ \setminus L$ and let $\mathbb{K}$ be the associated language of well-formed words. Observe that by definition, $\mathbb{K} \cup \mathbb{L}$ is the set of all well-formed words.

If $\mathbb{L}$ is definable in $\mathcal{F}$, then $\mathbb{L}$ is $\mathcal{F}$-separable from $\mathbb{K}$, hence by Theorem 4, L is $\mathcal{F}^+$-separable from K, and so L is definable in $\mathcal{F}^+$. Conversely, if L is definable in $\mathcal{F}^+$, then L is $\mathcal{F}^+$-separable from K and by Theorem 4, $\mathbb{L}$ is $\mathcal{F}$-separable from $\mathbb{K}$. Since $\mathbb{K} \cup \mathbb{L}$ is the set of all well-formed words, $\mathbb{L}$ is the intersection of the separator with the set of all well-formed words, which by hypothesis is also definable in $\mathcal{F}$. Therefore, $\mathbb{L}$ is definable in $\mathcal{F}$. ◀

Observe that being well-formed can be expressed in $\Pi_2(<)$: essentially, a word is well-formed if for all pairs of positions, either there is a third one in-between, or the labels of the two positions are "compatible". Hence, among the fragments of Figure 1, Theorem 6 applies to all fragments including and above $\Pi_2(<)$ in the quantifier alternation hierarchy. While such a transfer result was previously known [26, 16], the presentation and the proof are new. In particular, since membership is known to be decidable for $\Pi_2(<)$ [14], $\mathcal{B}\Sigma_2(<)$ [18] and $\Sigma_3(<)$ [18], we obtain new and simpler proofs of following results.

▶ **Corollary 7.** *Given a regular language L, one can decide whether*

- *L is definable by a $\Sigma_2(<,+1)$ (resp. $\Pi_2(<,+1)$) formula.*
- *L is definable by a $\mathcal{B}\Sigma_2(<,+1)$ formula.*
- *L is definable by a $\Sigma_3(<,+1)$ (resp. $\Pi_3(<,+1)$) formula.*

It remains to prove Theorem 4. We devote the rest of the section to this proof. An important remark is that the proof of the right to left direction is constructive: we start with an $\mathcal{F}$ formula that separates $\mathbb{L}$ from $\mathbb{L}'$ and use it to construct an $\mathcal{F}^+$ formula that separates L from L'. Note that the argument is generic for all fragments we consider.

On the other hand, the other direction, namely Proposition 9 below, requires a specific argument tailored to each fragment, which is a straightforward but tedious Ehrenfeucht-Fraïssé argument. Due to lack of space, we provide proofs of this proposition for each fragment in the long version of this paper.

4.1 From $\mathcal{F}^+$-separation to $\mathcal{F}$-separation

We prove that if L is $\mathcal{F}^+$-separable from L', then $\mathbb{L}$ is $\mathcal{F}$-separable from $\mathbb{L}'$. We actually prove the contrapositive: if $\mathbb{L}$ is *not* $\mathcal{F}$-separable from $\mathbb{L}'$, then L is *not* $\mathcal{F}^+$-separable from L'. We rely on a construction which, to any well-formed word $\mathfrak{u} \in \mathbb{A}_\alpha^+$ and any integer $i > 0$, associates a canonical word $\lceil \mathfrak{u} \rceil_i \in A^+$.

Canonical Word Associated to a Well-formed Word. To any $s \in S$, we associate an arbitrarily chosen nonempty word $\lceil s \rceil \in A^+$, such that $\alpha(\lceil s \rceil) = s$ (which is possible since α has been chosen surjective). Let $i > 0$. From a well-formed word $\mathfrak{u} \in \mathbb{A}_\alpha^+$, we build a word $\lceil \mathfrak{u} \rceil_i \in A^+$ as follows. If $\mathfrak{u} = s \in S$, then $\lceil \mathfrak{u} \rceil_i = \lceil s \rceil$ for all i. Otherwise, we have by definition

$$\mathfrak{u} = (s_0, e_1)(e_1, s_1, e_2) \cdots (e_{n-1} s_{n-1} e_n)(e_n, s_n).$$

For a natural $i > 0$, we set

$$\lceil \mathfrak{u} \rceil_i = \lceil s_0 \rceil \lceil e_1 \rceil^i \lceil s_1 \rceil \lceil e_2 \rceil^i \cdots \lceil e_{n-1} \rceil^i \lceil s_{n-1} \rceil \lceil e_n \rceil^i \lceil s_n \rceil .$$

Recall that β is the morphism $\beta : \mathbb{A}_\alpha^+ \to S$ mapping $\mathfrak{u}$ to $s_0 e_1 s_1 \cdots s_{n-1} e_n s_n$. Since $e_j \in E(S)$ for all j, it is immediate that $\alpha(\lceil \mathfrak{u} \rceil_i) = \beta(\mathfrak{u})$, hence we get the following fact:

▶ **Fact 8.** *For all $i > 0$ and all well-formed $\mathfrak{u} \in \mathbb{A}_\alpha^+$, we have $\mathfrak{u} \in \mathbb{L}$ (resp. $\in \mathbb{L}'$) if and only if $\lceil \mathfrak{u} \rceil_i \in L$ (resp $\in L'$).*

We now proceed with the proof. We use the classical preorders associated to fragments of first-order logic. The *(quantifier) rank* of a first-order formula φ is the largest number of quantifiers along a branch in the parse tree of φ. Given $u, v \in A^+$, we write $u \preceq_k^{+1} v$ if any $\mathcal{F}^+$ formula of rank k that is satisfied by u is satisfied by v as well. Similarly, for $\mathfrak{u}, \mathfrak{v} \in \mathbb{A}_\alpha^+$, we write $\mathfrak{u} \preceq_k \mathfrak{v}$ if any $\mathcal{F}$ formula of rank k that is satisfied by $\mathfrak{u}$ is satisfied by $\mathfrak{v}$ as well. One can verify that $\preceq_k$ and $\preceq_k^{+1}$ are preorders, as well as the following standard fact:

$$\begin{array}{l} L \subset A^+ \text{ is definable by an } \mathcal{F}^+\text{-formula of rank } k \text{ iff } L = \{u' \mid \exists u \in L \text{ st. } u \preceq_k^{+1} u'\} \\ \mathbb{L} \subset \mathbb{A}_\alpha^+ \text{ is definable by an } \mathcal{F}\text{-formula of rank } k \quad \text{iff } \mathbb{L} = \{\mathfrak{u}' \mid \exists \mathfrak{u} \in \mathbb{L} \text{ st. } \mathfrak{u} \preceq_k \mathfrak{u}'\}. \end{array} \tag{1}$$

Note that when $\mathcal{F}$ and $\mathcal{F}^+$ are closed under complement, then $\preceq_k$ and $\preceq_k^{+1}$ are actually equivalence relations. We can now state the main proposition of this direction:

▶ **Proposition 9.** *For any $k \in \mathbb{N}$, there exists $k' \in \mathbb{N}$ and $i \in \mathbb{N}$ such that for any well-formed words $\mathfrak{u}, \mathfrak{u}' \in \mathbb{A}_\alpha^+$ satisfying $\mathfrak{u} \preceq_{k'} \mathfrak{u}'$, we have $\lceil \mathfrak{u} \rceil_i \preceq_k^{+1} \lceil \mathfrak{u}' \rceil_i$.*

For all fragments of Figure 1, Proposition 9 is proved using classical Ehrenfeucht-Fraïssé arguments. While each proof is specific, the underlying ideas are similar. We present all these proofs in the long version of this paper. We finish the subsection by explaining how Proposition 9 can be used to terminate the proof of the first direction of Theorem 4.

Assume that $\mathbb{L}$ is *not* $\mathcal{F}$-separable from $\mathbb{L}'$. By definition this means that no language definable in $\mathcal{F}$ separates $\mathbb{L}$ from $\mathbb{L}'$. In particular for any k', the language $\{\mathfrak{u}' \mid \exists \mathfrak{u} \in \mathbb{L} \text{ st. } \mathfrak{u} \preceq_{k'} \mathfrak{u}'\}$ which is definable in $\mathcal{F}$ by (1), cannot be a separator. Note that this language contains $\mathbb{L}$. Hence, for all $k' \in \mathbb{N}$, there exist $\mathfrak{u} \in \mathbb{L}$ and $\mathfrak{u}' \in \mathbb{L}'$ such that $\mathfrak{u} \preceq_{k'} \mathfrak{u}'$. We deduce from Proposition 9 and Fact 8 that for all $k \in \mathbb{N}$, there exist $u \in L$ and $u' \in L'$ such that $u \preceq_k^{+1} u'$. It follows, again by (1), that L is *not* $\mathcal{F}^+$-separable from L', which terminates the proof.

4.2 From $\mathcal{F}$-separation to $\mathcal{F}^+$-separation

We now prove that if $\mathbb{L}$ is $\mathcal{F}$-separable from $\mathbb{L}'$, then L is $\mathcal{F}^+$-separable from L', by building an $\mathcal{F}^+$-definable separator. We rely on a construction that is dual to the one used previously: to any word $w \in A^+$, we associate a canonical well-formed word $\lfloor w \rfloor \in \mathbb{A}_\alpha^+$.

Canonical Well-formed Word Associated to a Word. To any word w of A^+, we associate a canonical well-formed word $\lfloor w \rfloor \in \mathbb{A}_\alpha^+$ such that $\alpha(w) = \beta(\lfloor w \rfloor)$. This construction is adapted from [20] and is originally inspired by [26]. Fix an arbitrary order on the set $E(S)$.

For a position x of w, let $u_x \in A^+$ be the infix of w obtained by keeping only positions $x - (|S| - 1)$ to x. If position $x - (|S| - 1)$ does not exist, u_x is just the prefix of w ending at x. A position x is said *distinguished* if there exists an idempotent $e \in E(S)$ such that $\alpha(u_x) \cdot e = \alpha(u_x)$. Additionally, we always define the rightmost position as distinguished, even if it does not satisfy the property. Set $x_1 < \cdots < x_{n+1}$ as the distinguished positions in w, so that x_{n+1} is the rightmost position. Let $e_1, \ldots, e_n \in E(S)$ be such that for all i, e_i is the smallest idempotent such that $\alpha(u_{x_i}) \cdot e_i = \alpha(u_{x_i})$.

If $n = 0$, *i.e.*, if the only distinguished position is the rightmost one, set $\lfloor w \rfloor = \alpha(w) \in \mathbb{A}_\alpha$. Otherwise, we define $\lfloor w \rfloor \in \mathbb{A}_\alpha^+$ as the word:

$$\lfloor w \rfloor = (\alpha(w_0), e_1) \cdot (e_1, \alpha(w_1), e_2) \cdots (e_{n-1}, \alpha(w_{n-1}), e_n) \cdot (e_n, \alpha(w_n)) \tag{2}$$

where w_0 is the prefix of w ending at position x_1, for all $1 \leqslant i \leqslant n-1$, w_i is the infix of w obtained by keeping positions $x_i + 1$ to x_{i+1}, and w_n is the suffix of w starting at position $x_n + 1$. Note that $\lfloor w \rfloor$ is well-formed. The next fact follows from the definitions.

▶ **Fact 10.** *For all $w \in A^+$, $\alpha(w) = \beta(\lfloor w \rfloor)$. Hence, $w \in L$ iff $\lfloor w \rfloor \in \mathbb{L}$ and $w \in L'$ iff $\in \mathbb{L}'$.*

To any distinguished position x_i in w, we associate position $\lfloor x \rfloor = i$ in $\lfloor w \rfloor$. Our main motivation for using this construction is its local canonicity, stated in the following lemma.

▶ **Lemma 11.** *Let $w \in A^+$. Then we have the following properties:*

(*a*) *whether a position x is distinguished in w, and if so the label of position $\lfloor x \rfloor$ in $\lfloor w \rfloor$ only depends on the infix of w of length $2|S|$ ending at position x. That is, if the infixes of length $2|S|$ ending at x and y are equal, then x is distinguished iff so is y, and in that case, the labels of $\lfloor x \rfloor$ and $\lfloor y \rfloor$ in $\lfloor w \rfloor$ are equal.*

(*b*) *the label of the last position of $\lfloor w \rfloor$ only depends on the suffix of length $2|S|$ of w.*

Proof. It is immediate that whether x is distinguished and if so the associated idempotent only depends on the infix u_x of length at most $|S|$ ending at x. Therefore, to prove (*a*), it suffices to show that all infixes w_i in (2) are of size at most $|S|$, or in other words, that among $|S| + 1$ consecutive positions, at least one is distinguished. So let us consider an infix $a_1 \cdots a_{|S|+1}$ of w of length $|S| + 1$. It is immediate from the pigeonhole principle that there exist $i < j$ such that $\alpha(a_1 \cdots a_i) = \alpha(a_1 \cdots a_j) = \alpha(a_1 \cdots a_i) \cdot (\alpha(a_{i+1} \cdots a_j))^\omega$. Hence, the position corresponding to a_i is distinguished. The proof of the second assertion is similar. ◀

L is $\mathcal{F}^+$-separable from L'. We can now construct our separator. The construction follows from the next proposition.

▶ **Proposition 12.** *Let $\mathbb{K} \subseteq \mathbb{A}_\alpha^+$ that can be defined using an $\mathcal{F}$ formula φ. Then there exists an $\mathcal{F}^+$ formula Ψ over alphabet A such that for every word $w \in A^+$:*

$$w \models \Psi \text{ if and only if } \lfloor w \rfloor \models \varphi.$$

Proof. Proposition 12 follows from the following simple consequence of Lemma 11.

▶ **Claim 13.** *For any* $\mathtt{a} \in \mathbb{A}_\alpha$ *there exists a formula* $\gamma_{\mathtt{a}}(x)$ *of* $\mathcal{F}^+$ *with a free variable* x*, such that for any* $w \in A^+$ *and any position* x *of* w*, we have* $w \models \gamma_{\mathtt{a}}(x)$ *iff* x *is distinguished and* $\lfloor x \rfloor$ *has label* $\mathtt{a}$ *in* $\lfloor w \rfloor$*.*

This claim holds since by Lemma 11, formula $\gamma_{\mathtt{a}}(x)$ only needs to explore the neighborhood of size $2|S|$ of x, which is trivially possible for all fragments $\mathcal{F}^+$ we consider.

To conclude the proof of Proposition 12, it suffices to define Ψ as the formula constructed from φ by restricting all quantifiers to positions that are distinguished and to replace all tests $P_{\mathtt{a}}(x)$ by $\gamma_{\mathtt{a}}(x)$. ◀

We can now finish the proof of Theorem 4. Assume that $\mathbb{L}$ is $\mathcal{F}$-separable from $\mathbb{L}'$ and let φ be an $\mathcal{F}$ formula defining a separator. We denote by Ψ the $\mathcal{F}^+$ formula obtained from φ as defined in Proposition 12. We prove that Ψ defines a language separating L from L'.

We first prove that $L \subseteq \{w \mid w \models \Psi\}$. When $w \in L$, by Fact 10, we have $\lfloor w \rfloor \in \mathbb{L}$. Hence, $\lfloor w \rfloor \models \varphi$ and so $w \models \Psi$ by definition of Ψ. The proof that $L' \subseteq \{w \mid w \not\models \Psi\}$ is identical: if $w \in L'$, we have $\lfloor w \rfloor \in \mathbb{L}'$ by Fact 10. Hence, $\lfloor w \rfloor \not\models \varphi$ and $w \not\models \Psi$ by definition of Ψ. ◀

5 Algebraic Approach

We now present an algebraic version of Theorem 4: the operator $\mathsf{V} \mapsto \mathsf{V} * \mathsf{D}$ preserves decidability of separation. We would like to emphasize again that the ideas behind this theorem are essentially the same as for Theorem 4. In particular, proofs presented in the long version of this paper only rely on elementary notions, thus bypassing complex constructions usually used to prove this kind of result, even if the statement itself requires some additional algebraic vocabulary.

The section is organized in three parts. We first briefly recall how classes of languages corresponding to our logical fragments are given an algebraic definition: for each fragment, an associated class of finite semigroups (or monoids) V, a *variety*, has already been characterized, such that the class of languages definable in the fragment is exactly the class of languages that are recognized by a semigroup (or monoid) of V. In the second part, we define what "adding the successor relation" means in this context. Given a variety V, this generally corresponds to considering a new variety built on top of V via an operation called the *semidirect product*. This new variety is denoted $\mathsf{V} * \mathsf{D}$. Finally, in the last part, we state our main theorem: for any variety V, separability for the variety $\mathsf{V} * \mathsf{D}$ reduces to separability for the variety V.

5.1 Varieties

A *variety of semigroups (resp. monoids)* is a class of finite semigroups (resp. monoids) closed under three natural operations: finite direct product, subsemigroup (or submonoid), and homomorphic image. A variety V defines a class of languages, also noted V, namely the class of all of languages recognized by semigroups (resp. monoids) in V. There is an issue however: all classes of languages defined in this way have to be closed under complement, since the set of languages recognized by any semigroup is closed under complement. This prevents us from capturing logical fragments that are not closed under complement, such as $\Sigma_2(<)$. This problem has been solved in [13] with the notions of *ordered semigroups and monoids*. Intuitively, such a semigroup is parametrized by a partial order and the set of languages it recognizes is then restricted with respect to this partial order. These classical constructions

will be recalled in the long version of this paper, as well as varieties corresponding to all fragments we deal with.

Logical fragments presented in Sec. 2 correspond to varieties that have been fully identified. For each fragment, its non-enriched variant corresponds to a variety V of (ordered) monoids and its enriched version to the variety of (ordered) semigroups $\mathsf{V} * \mathsf{D}$ built from V. For example, $\mathrm{FO}^2(<)$ corresponds to the variety of monoids DA and $\mathrm{FO}^2(<, +1)$ to the variety of semigroups $\mathsf{DA} * \mathsf{D}$ [29] (see the long version for a bibliography with all correspondences).

5.2 Semidirect Product

The Variety D. The variety D consists of all finite ordered semigroups S such that for all $s \in S$ and all $e \in E(S)$, we have $se = e$. From a language perspective, a language L is recognized by a semigroup in D iff there exists $k \in \mathbb{N}$ such that membership of a word w to L only depends on the suffix of length k of w.

Semidirect Product. Let M be an ordered monoid and T be an ordered semigroup. A *semidirect product* of M and T is an operation which is parametrized by an *action* of T on M and outputs a new ordered semigroup, whose base set is $M \times T$. Therefore, one can obtain different semidirect products out of the same M and T, depending on the chosen action (we recall the construction in the long version). One can next lift this product at the level of varieties.

We are interested in the semidirect products of the form $\mathsf{V} * \mathsf{D}$, the variety of ordered semigroups generated by all semidirect products of an ordered monoid of V by an ordered semigroup of D. The reason why we introduce such semidirect products is the following theorem, which gathers several nontrivial results from the literature (see the long version).

▶ **Theorem 14.** *Let V be a variety corresponding to a fragment $\mathcal{F}$ from the ones presented in Figure 1. Then, the variety corresponding to the fragment $\mathcal{F}^+$ is $\mathsf{V} * \mathsf{D}$.*

5.3 Main Theorem

We have now the machinery needed to state our main theorem. For any variety of ordered monoids V, we reduce $(\mathsf{V} * \mathsf{D})$-separability to V-separability.

▶ **Theorem 15.** *Let V be a non-trivial variety of ordered monoids. Let L and L' be two languages both recognized by the same morphism $\alpha : A^+ \to S$ into a finite semigroup S. Set $\mathbb{L}, \mathbb{L}' \subseteq \mathbb{A}_\alpha^+$ as the languages of well-formed words associated to L, L'. Then L is $(\mathsf{V} * \mathsf{D})$-separable from L' if and only if $\mathbb{L}$ is V-separable from $\mathbb{L}'$.*

The proof of Theorem 15 is presented in the long version of this paper. As it was the case for Theorem 4, the proof is both elementary and constructive: if there exists a separator for $\mathbb{L}$ and $\mathbb{L}'$ in V, we use it to construct a separator for L and L' in $\mathsf{V} * \mathsf{D}$.

In view of Theorem 14, Theorem 15 applies to all fragments we introduced. This means that Theorem 4 can be given an alternate indirect proof within this algebraic framework by combining Theorem 15 and Theorem 14. Hence, this also yields another proof of Corollary 5.

6 Conclusion

We proved that separation is decidable over finite words for the following logical fragments: $\mathrm{FO}(=, +1)$, $\Sigma_1(<, +1, min, max)$, $\mathcal{B}\Sigma_1(<, +1, min, max)$, $\Sigma_2(<, +1)$ and $\mathrm{FO}^2(<, +1)$. To

achieve this, we presented a simple reduction to the same problem for the weaker fragments $\mathrm{FO}(=)$, $\Sigma_1(<)$, $\mathcal{B}\Sigma_1(<)$, $\Sigma_2(<)$ and $\mathrm{FO}^2(<)$.

The reduction itself is entirely generic to all fragments and its proof is elementary, and also mostly generic. In particular, the technique can be used to prove that the reduction works for other natural fragments of first-order logic. An interesting example to which these results apply is the quantifier alternation hierarchy within $\mathrm{FO}^2(<)$ (known as the Trotter-Weil hierarchy, and which is decidable [11]). However, the separation problem for classes in this hierarchy has yet to be investigated. We also obtained direct proofs that membership is decidable for $\mathcal{B}\Sigma_2(<,+1)$ and $\Sigma_3(<,+1)$.

Finally, we presented an algebraic formulation of this reduction, which recovers a previously known result by Steinberg [24], while having a much simpler proof. One can expect extending these results to other fragments, such as enrichment with modulo predicates. Another advantage of this technique is that it can be extended in a straightforward way to the same logical fragments over words of infinite length. This yields identical transfer results. We leave the presentation of these results for further work.

References

1 J. Almeida. A syntactical proof of locality of DA. *Internat. J. Algebra Comput.*, 6, 1996.
2 Jorge Almeida. Some algorithmic problems for pseudovarieties. *Publ. Math. Debrecen*, 54, 1999. Proc. of Automata and Formal Languages, VIII.
3 Karl Auinger. On the decidability of membership in the global of a monoid pseudovariety. *IJAC*, 20(2), 2010.
4 Rina S. Cohen and J.A. Brzozowski. Dot-depth of star-free events. *J. Comp. Syst. Sci.*, 5, 1971.
5 Wojciech Czerwiński, Wim Martens, and Tomáš Masopust. Efficient separability of regular languages by subsequences and suffixes. In *ICALP'13*, 2013.
6 Volker Diekert and Paul Gastin. First-order definable languages. In *Logic and Automata: History and Perspectives*, volume 2. Amsterdam Univ. Press, 2008.
7 Christian Glaßer and Heinz Schmitz. Languages of dot-depth 3/2. *Theory Comp. Syst.*, 42(2), 2008.
8 Karsten Henckell. Pointlike sets: the finest aperiodic cover of a finite semigroup. *J. Pure Appl. Algebra*, 55(1-2), 1988.
9 Robert Knast. A semigroup characterization of dot-depth one languages. *Rairo* ITA, 17(4), 1983.
10 Manfred Kufleitner and Alexander Lauser. Around dot-depth 1. *Int. J. Found. Comp. Sci.*, 23(6), 2012.
11 Manfred Kufleitner and Pascal Weil. On logical hierarchies within FO^2-definable languages. *Logical Methods in Computer Science*, 8(3), 2012.
12 Robert McNaughton and Seymour Papert. *Counter-Free Automata.* MIT Press, 1971.
13 Jean-Eric Pin. A variety theorem without complementation. *Russian Mathematics*, 39, 1995.
14 Jean-Éric. Pin and Pascal Weil. Polynomial closure and unambiguous product. *Theory Comp. Syst.*, 30(4), 1997.
15 Jean-Eric Pin and Pascal Weil. Semidirect products of ordered semigroups. *Comm. Alg.*, 30, 2002.
16 Jean-Eric Pin and Pascal Weil. The wreath product principle for ordered semigroups. *Communications in Algebra*, 30:5677–5713, 2002.
17 T. Place, L. van Rooijen, and M. Zeitoun. Separating regular languages by locally testable and locally threshold testable languages. In *FSTTCS'13*, LIPIcs, 2013.

18 T. Place and M. Zeitoun. Going higher in the first-order quantifier alternation hierarchy on words. In *ICALP'14*, volume 8573 of *LNCS*, 2014. `http://arxiv.org/pdf/1404.6832v1`.

19 T. Place and M. Zeitoun. Separating regular languages with FO. In *CSL-LICS'14*, 2014.

20 Thomas Place and Luc Segoufin. Decidable characterization of $\mathrm{FO}^2(<, +1)$ and locality of DA. Unpublished, to appear, 2014.

21 Thomas Place, Lorijn van Rooijen, and Marc Zeitoun. Separating regular languages by piecewise testable and unambiguous languages. In *MFCS' 13*, 2013.

22 Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *I. Control*, 8, 1965.

23 I. Simon. Piecewise testable events. In *Automata Theory and Formal Languages 2nd GI Conference*, volume 33 of *LNCS*. Springer, 1975.

24 Benjamin Steinberg. A delay theorem for pointlikes. *Sem. Forum*, 63(3), 2001.

25 Howard Straubing. A generalization of the Schützenberger product of finite monoids. *TCS*, 1981.

26 Howard Straubing. Finite semigroup varieties of the form $\mathsf{V}*\mathsf{D}$. *J. Pure Appl. Algebra*, 36, 1985.

27 Denis Thérien. Classification of finite monoids: the language approach. *TCS*, 4(2), 1981.

28 Denis Thérien and Alex Weiss. Graph congruences and wreath products. *J. Pure Appl. Algebra*, 36, 1985.

29 Denis Thérien and Thomas Wilke. Over words, two variables are as powerful as one quantifier alternation. In *STOC' 98*. ACM, 1998.

30 Wolfgang Thomas. Classifying regular events in symbolic logic. *J. Comp. Syst. Sci.*, 25(3), 1982.

A Varieties and Logical Fragments

In this appendix, we look at each logical fragment that we consider in the paper and present a state of the art.

We proceed as follows. First we define ordered semigroups and monoids. Then we give the definition of varieties of finite ordered semigroups. Note that for technical reasons we will need to consider both varieties of semigroups and monoids (non-enriched fragments correspond to varieties of monoids while enriched ones correspond to varieties of semigroups). For the sake of simplifying the presentation, we only give the definitions for semigroups. Ordered monoids and varieties of ordered monoids are defined in a similar way. One should actually have in mind that there is:

1. a generic theorem, called Eilenberg's theorem, which establishes a correspondence between varieties and classes of languages (indexed by alphabets) having certain closure properties. It was first obtained by S. Eilenberg for classes closed under complement, and generalized by J.E Pin [13] when this assumption does not necessarily hold. We shall not state this theorem.
2. specific theorems, one for each class, relating a class of languages with a corresponding variety of ordered semigroups or monoids. We will recall these connections for all logically defined fragments presented in Section 2.

Ordered Semigroups. An ordered semigroup is a pair $(S, \leqslant)$ where S is a semigroup and $\leqslant$ is a partial order on S, which is compatible with multiplication: $s \leqslant t$ and $s' \leqslant t'$ imply $ss' \leqslant tt'$. To simplify the notation, we will often omit the partial order $\leqslant$ when it is clear from the context and simply speak of an ordered semigroup S. Observe that any semigroup endowed with equality as the partial order is an ordered semigroup. In particular we view A^+ as an ordered semigroup with equality as the partial order.

If $(S, \leqslant_S)$ and $(T, \leqslant_T)$ are ordered semigroups, an ordered semigroup morphism is a mapping $\alpha : S \to T$ which is a semigroup morphism and preserves the partial order, *i.e.*, for all $s, s' \in S$, $s \leqslant_S s' \Rightarrow \alpha(s) \leqslant_T \alpha(s')$. Let $L \subseteq A^+$ and $(S, \leqslant)$ be an ordered semigroup. Then, L is said to be *recognized by* $(S, \leqslant)$ if there exist an ordered semigroup morphism $\alpha : A^+ \to S$ and $F \subseteq S$, such that $L = \alpha^{-1}(F)$ and F is *upward closed*, that is:

$$s \in F \text{ and } s \leqslant t \Rightarrow t \in F.$$

When $\leqslant$ is trivial, this is exactly the classical notion of recognizability by semigroups (see above). However, when $\leqslant$ is nontrivial, the set of recognized languages gets restricted because of the additional condition on the recognizing set F. In particular it may happen that a language is recognized by $(S, \leqslant)$, while its complement is not.

Varieties of Ordered Semigroups. A *variety of finite ordered semigroups* is a class V of finite ordered semigroups that satisfies the following properties:

1. V is closed under ordered subsemigroup: if $(S, \leqslant) \in \mathsf{V}$, then $(T, \leqslant) \in \mathsf{V}$ when T is a subsemigroup of S and the order on T is the restriction of the order on S.
2. V is closed under ordered quotient: if $(S, \leqslant) \in \mathsf{V}$ and $\alpha : (S, \leqslant) \to (T, \leqslant)$ is a surjective ordered semigroup morphism, then $(T, \leqslant) \in \mathsf{V}$.
3. V is closed under Cartesian direct product: if $(S_1, \leqslant_1), (S_2, \leqslant_2) \in \mathsf{V}$, then $(S_1 \times S_2, \leqslant) \in \mathsf{V}$, where the semigroup $S_1 \times S_2$ is equipped with the componentwise multiplication and $(s_1, s_2) \leqslant (t_1, t_2)$ if $s_1 \leqslant_1 t_1$ and $s_2 \leqslant_2 t_2$.

Varieties of finite ordered monoids, of finite semigroups and of finite monoids are defined analogously.

As explained above, we use varieties to define classes of languages. To a variety V, we associate the class of all languages that can be recognized by an ordered semigroup (resp. ordered monoid) in V. As for logics, for the sake of simplifying the presentation, we may abuse notation and use V to denote both a variety and the class of languages it defines.

Note that the bibliography presented here is needed to formally prove Theorem 14 and therefore for obtaining an alternate proof of Theorem 4 in the algebraic framework of Section 5. Namely, for each fragment $\mathcal{F}$ in Figure 1, we present references solving the following questions:

1. Connecting the weak variant $\mathcal{F}$ to a variety of monoids V.
2. Connecting the strong variant $\mathcal{F}^+$ to the variety of semigroups $\mathsf{V} * \mathsf{D}$.
3. Solving the separation problem for $\mathcal{F}$.

First-order with Equality. FO(=) is the restriction of FO(<) in which the linear order cannot be used, and only equality between two positions can be tested. It is folklore that FO(=)-definable languages are exactly those that can be defined using a monoid in the variety of monoids ACom of aperiodic and commutative monoids. The enriched fragment FO(=, +1) (*min* and *max* can be defined in the logic) defines locally threshold testable languages [30]. In [28], it was proven that FO(=, +1)-definable languages are exactly those that can be defined in $\mathsf{ACom} * \mathsf{D}$. In particular this was used to solve the membership problem for FO(=, +1).

That separation is decidable for FO(=) is simple (essentially the problem can be reduced to the decision of Presburger logic, see [17]). Hence Theorem 4 and Theorem 15 yield two different proofs of the following corollary.

▶ **Corollary 16.** *Let L, L' be regular languages. It is decidable to test whether L is* FO(=, +1)*-separable from L'.*

As we explained in Section 2, while the proof of Corollary 16 is new, the result itself is not. A specific proof was presented in [17] and the result can also be obtained through indirect means by combining results from [2, 24].

Quantifier Alternation Hierarchy. One can classify first-order formulas by counting the number of alternations between $\exists$ and $\forall$ quantifiers in the prenex normal form of the formula. Set $i \in \mathbb{N}$, a formula is said to be $\Sigma_i(<)$ (resp. $\Pi_i(<)$) if its prenex normal form has $(i-1)$ quantifier alternations (*i.e.*, i blocks of quantifiers) and starts with an $\exists$ (resp. $\forall$) quantifier. For example, a formula whose prenex normal form is

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

is $\Sigma_3(<)$. Observe that a $\Pi_i(<)$ formula is by definition the negation of a $\Sigma_i(<)$ formula. Finally, a $\mathcal{B}\Sigma_i(<)$ formula is a boolean combination of $\Sigma_i(<)$ formulas.

Both this hierarchy and the enriched variant are known to be strict. In particular, they correspond to well-known hierarchies of classes of languages. The non-enriched hierarchy corresponds to the Straubing-Thérien hierarchy [25, 27]. The enriched hierarchy corresponds to the dot-depth hierarchy [4]. Note that for all fragments above $\Sigma_2(<)$, the predicates *min* and *max* can be expressed in the logic. Hence, we denote the enriched fragments by $\Sigma_1(<, +1, min, max)$, $\mathcal{B}\Sigma_1(<, +1, min, max)$, $\Sigma_2(<, +1), \ldots$

Solving the membership problem for all levels in both hierarchies has been an open problem for a long time. As of today, only the lower levels are known to be decidable.

Historically, $\mathcal{B}\Sigma_1(<)$ and $\mathcal{B}\Sigma_1(<, +1, min, max)$ have been investigated first. It is known from [23] that $\mathcal{B}\Sigma_1(<)$ has decidable membership and corresponds to the variety of monoids J. For $\mathcal{B}\Sigma_1(<, +1, min, max)$, decidability was proven in [9], as well as the correspondence with the variety of semigroups $\mathsf{J} * \mathsf{D}$ in [26].

The fragments $\Sigma_1(<)$ and $\Sigma_2(<)$ were proved to have decidable membership in [14]. Moreover, the authors also prove that each of these two fragments correspond to varieties of ordered monoids and that $\Sigma_1(<, +1, min, max)$ and $\Sigma_2(<, +1)$ correspond to the varieties of semigroups obtained by taking the semidirect product with D. From this correspondence, they obtain decidability of $\Sigma_1(<, +1, min, max)$. This is more involved for $\Sigma_2(<, +1)$ and was proven later in [7].

Recently, membership has been shown to be decidable for both $\mathcal{B}\Sigma_2(<)$ and $\Sigma_3(<)$ [18]. These results can be transferred to $\mathcal{B}\Sigma_2(<, +1)$ and $\Sigma_3(<, +1)$ using a result by Straubing [26], or Theorem 6 in this paper. For all levels above, the membership problem is open.

Separation is known to be decidable for $\Sigma_1(<)$ [5], $\mathcal{B}\Sigma_1(<)$ [21, 5] and $\Sigma_2(<)$ [18]. Hence Theorem 4 and Theorem 15 yield two different proofs of the following corollary.

▶ **Corollary 17.** *Let L, L' be regular languages, then the following problems are decidable:*

- *whether L is $\Sigma_1(<, +1, min, max)$-separable from L'.*
- *whether L is $\mathcal{B}\Sigma_1(<, +1, min, max)$-separable from L'.*
- *whether L is $\Sigma_2(<, +1)$-separable from L'.*

As we explained in Section 2, the result for $\mathcal{B}\Sigma_1(<, +1, min, max)$ as it can also be obtained through indirect means by combining results from [2, 24]. On the other hand, the results are new for both $\Sigma_1(<, +1, min, max)$ and $\Sigma_2(<, +1)$.

Two-Variable First-Order Logic. $\mathrm{FO}^2(<)$ is the restriction of $\mathrm{FO}(<)$ using only two (reusable) variables. The corresponding enriched fragment is $\mathrm{FO}^2(<, +1)$ (*min* and *max* can be expressed in the logic).

In [29], it was proven that $\mathrm{FO}^2(<)$ and $\mathrm{FO}^2(<, +1)$ correspond respectively to the varieties DA and $\mathsf{DA} * \mathsf{D}$. This immediately yields decidability of membership for $\mathrm{FO}^2(<)$. For $\mathrm{FO}^2(<, +1)$, this additionally requires a deep algebraic result by Almeida [1] (a simpler self-contained proof also exists [20]). The separation problem has been proved to be decidable for $\mathrm{FO}^2(<)$ in [21]. Hence Theorem 4 and Theorem 15 yield two different proofs of the following corollary.

▶ **Corollary 18.** *Let L, L' be regular languages. It is decidable to test whether L is $\mathrm{FO}^2(<, +1)$-separable from L'.*

As we explained in Section 2, while the proof is new, the result itself is not. It can also be obtained through indirect means, again by combining results from [2, 24].

B Appendix to Section 4: Proof of Proposition 9

In this appendix, we prove Proposition 9. As explained in the main paper, this requires a specific proof for each fragment we consider. Here, we consider two main cases, $\mathcal{F} = \mathrm{FO}^2(<)$ and $\mathcal{F} = \Sigma_n(<)$ for some n. Note that we will obtain the case $\mathcal{F} = \mathcal{B}\Sigma_n(<)$ as a simple consequence of the $\Sigma_n(<)$ case. Finally, we leave out the case $\mathcal{F} = \mathrm{FO}(=)$, as the argument is essentially a copy and paste of the $\Sigma_n(<)$ argument.

B.1 $\mathbf{FO}^2(<)$ and $\mathbf{FO}^2(<,+1)$

Observe that since $\mathrm{FO}^2(<)$ and $\mathrm{FO}^2(<,+1)$ are both closed under complement, the preorders $\preceq_k$ and $\preceq_k^{+1}$ are actually equivalence relations. To avoid confusion with other fragments, we denote by $\equiv_k$ and $\equiv_k^{+1}$, these two equivalences. We prove the following proposition, which clearly entails Proposition 9.

▶ **Proposition 19.** *For any $k \in \mathbb{N}$, given $\mathfrak{u}, \mathfrak{u}' \in \mathbb{A}_\alpha^+$ we have the following implication:*

$$\mathfrak{u} \equiv_k \mathfrak{u}' \Rightarrow \lceil \mathfrak{u} \rceil_{2k} \equiv_k^{+1} \lceil \mathfrak{u}' \rceil_{2k}$$

This is proved using an Ehrenfeucht-Fraïssé argument. We first define the Ehrenfeucht-Fraïssé game associated to $\mathrm{FO}^2(<)$ (*i.e.,* corresponding to $\equiv_k$) and then explain how to adapt it to $\equiv_k^{+1}$.

Ehrenfeucht-Fraïssé Game. The board of the $\mathrm{FO}^2(<)$-game consists of two words and lasts a predefined number k of rounds. There are two players called *Spoiler and Duplicator*. At any time during the game there is one pebble placed on a position of one word and one pebble placed on a position of the other word, and both positions have the same label. When the game starts, both pebbles are placed on the first position of each words. Each round starts with Spoiler choosing one of the pebbles, and moving it inside its word from its original position x to a new position y. Duplicator must answer by moving the other pebble in the other word from its original position x' to a new position y'. Moreover, x' and y' must satisfy the same relations as x and y among '$<$' and the label predicates.

Duplicator wins if she manages to play for all k rounds. Spoiler wins as soon as Duplicator is unable to play.

The $\mathrm{FO}^2(<,+1)$-game is defined similarly with additional constraints for Duplicator. When Spoiler makes a move, Duplicator's must choose is answer y' so that x' and y' satisfy the same relations as x and y among $+1$, $<$ and the label predicates.

▶ **Lemma 20** (Folklore). *For any integer k and any words v, v', we have the following facts:*

- *$v \equiv_k v'$ iff Duplicator has a winning strategy in the k-round $\mathrm{FO}^2(<)$-game on v and v'.*
- *$v \equiv_k^{+1} v'$ iff Duplicator has a winning strategy in the k-round $\mathrm{FO}^2(<,+1)$-game on v and v'.*

Set $u = \lceil \mathfrak{u} \rceil_{2k}$ and $u' = \lceil \mathfrak{u}' \rceil_{2k}$, with the notation of Lemma 20. In view of Lemma 20, we may prove that $u \equiv_k^{+1} u'$ by giving a winning strategy in the k-round $\mathrm{FO}^2(<,+1)$-game played on u and u'. We call $\mathcal{G}$ this game. The strategy involves playing a shadow $\mathrm{FO}^2(<)$-game $\mathcal{S}$ on $\mathfrak{u}$ and $\mathfrak{u}'$. Observe that by hypothesis and Lemma 20, Duplicator has a winning strategy for k rounds in the game $\mathcal{S}$. We begin by setting up some notation to help us define Duplicator's strategy.

Notation. Assuming that $\mathfrak{u} \preceq_k \mathfrak{u}'$, we need to prove that $u \preceq_k^{+1} u'$. If $\mathfrak{u} \in S$ or $\mathfrak{u}' \in S$, then $\mathfrak{u} = \mathfrak{u}' = s \in S$ (since the only well-formed word that contains letter $s \in S$ is s itself) and the result is immediate.

Otherwise, by hypothesis, the words $\mathfrak{u}$ and $\mathfrak{u}'$ are of the form

$$\begin{aligned} \mathfrak{u} &= (s_0, e_1)(e_1, s_1, e_2) \cdots (e_m, s_m) \\ \mathfrak{u}' &= (s'_0, e'_1)(e'_1, s'_1, e'_2) \cdots (e'_{m'}, s'_{m'}) \end{aligned}$$

In particular, observe that since $\mathfrak{u} \preceq_k \mathfrak{u}'$ and the labels of the leftmost and rightmost positions are unique in $\mathfrak{u}$ and $\mathfrak{u}'$, we have $s_0 = s'_0$, $e_1 = e'_1$, $e_m = e'_{m'}$ and $s_m = s'_{m'}$. For the

sake of simplifying the presentation, we assume that for all $i \leqslant m$, we have $\lceil s_i \rceil = a_i \in A$ and $\lceil e_i \rceil = b_i \in A$ (this does not harm the generality of the proof). Similarly, for all $i \leqslant m'$, we assume that $\lceil s'_i \rceil = a'_i \in A$ and $\lceil e'_i \rceil = b'_i \in A$. By definition, we have

$$\begin{array}{lcl} u & = & \lceil \mathfrak{u} \rceil_{2k} = a_0(b_1)^{2k}a_1(b_2)^{2k}\cdots(b_m)^{2k}a_m \\ u' & = & \lceil \mathfrak{u}' \rceil_{2k} = a'_0(b'_1)^{2k}a'_1(b'_2)^{2k}\cdots(b'_{m'})^{2k}a'_{m'} \end{array}$$

Winning Strategy. We define an invariant $\mathcal{I}(\ell)$ (ℓ is the number of remaining rounds) that Duplicator has to satisfy when playing. Assume that the pebbles in u, u' are at positions x, x' in $\mathcal{G}$ and that the pebbles in $\mathfrak{u}, \mathfrak{u}'$ are at positions i, i' in $\mathcal{S}$. Then, $\mathcal{I}(\ell)$ holds when so do all following properties:

1. Duplicator has a winning strategy for playing ℓ rounds in $\mathcal{S}$. In particular this means that i, i' have the same label and that $b_i = b'_{i'}$, $a_i = a'_{i'}$ and $b_{i+1} = b'_{i'+1}$.
2. x and x' are inside the identical factors $(b_i)^{2k}a_i(b_{i+1})^{2k}$ and $(b'_{i'})^{2k}a'_{i'}(b'_{i'+1})^{2k}$, and at the same relative position.
3. there are at least ℓ copies of b_i (resp $b'_{i'}$) to the left of x (resp. x') and ℓ copies of b_{i+1} (resp. $b'_{i'+1}$) to the right of x resp. x').

It is clear that $\mathcal{I}(k)$ holds at the beggining of the game. Assume now that $\mathcal{I}(\ell+1)$ holds and that there are $\ell+1$ rounds left to play. We explain how Duplicator can answer a move by Spoiler while enforcing $\mathcal{I}(\ell)$. Assume that Spoiler moves the pebble in u to a new position y (the dual case, when Spoiler plays in u', is treated similarly). There are two distinct cases.

- If y remains in the factor $(b_i)^{2k}a_i(b_{i+1})^{2k}$ and satifies Item 3 of $\mathcal{I}(\ell)$, then Duplicator simply copies Spoiler's move in $(b'_{i'})^{2k}a'_{i'}(b'_{i'+1})^{2k}$. The positions i and i' remain unchanged and $\mathcal{I}(\ell)$ is clearly satisfied.
- Otherwise, let $j \neq i$ such that x belongs to $(b_j)^{2k}a_j(b_{j+1})^{2k}$, has at least ℓ copies of b_j to its left and ℓ copies of b_{j+1} to its right. To get an answer, Duplicator simulates a move by Spoiler in $\mathcal{S}$ by moving the pebble from position i to position j. From her winning strategy in $\mathcal{S}$, she obtains a position j' in $\mathfrak{u}'$. This gives her a position y' in $(b'_{j'})^{2k}a'_{j'}(b'_{j'+1})^{2k}$ which satisfies $\mathcal{I}(\ell)$. This terminates the proof.

B.2 $\Sigma_n(<)$ and $\Sigma_n(<, +1, min, max)$

We fix some $n \in \mathbb{N}$. We keep using the symbols $\preceq_k$ and $\preceq_k^{+1}$ to denote the preorders associated to $\Sigma_n(<)$ and $\Sigma_n(<, +1, min, max)$. Furthermore, we denote by $\cong_k$ and $\cong_k^{+1}$ the equivalence relations associated to $\mathcal{B}\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<, +1, min, max)$. We prove the following proposition, which again yields Proposition 9 with $k' = k$ and $i = 2^{k+1}$.

▶ **Proposition 21.** *For any $k \in \mathbb{N}$, given $\mathfrak{u}, \mathfrak{u}' \in \mathbb{A}_\alpha^+$ we have the following implications:*

$$\begin{array}{lcl} \mathfrak{u} \preceq_k \mathfrak{u}' & \Rightarrow & \lceil \mathfrak{u} \rceil_{2^{k+1}} \preceq_k^{+1} \lceil \mathfrak{u}' \rceil_{2^{k+1}} \\ \mathfrak{u} \cong_k \mathfrak{u}' & \Rightarrow & \lceil \mathfrak{u} \rceil_{2^{k+1}} \cong_k^{+1} \lceil \mathfrak{u}' \rceil_{2^{k+1}} . \end{array}$$

Observe first that the second implication is an immediate consequence of the first one. Indeed, since $\mathcal{B}\Sigma_n$ formulas are boolean combinations of Σ_n formulas, we have

$$\begin{array}{rcl} v \preceq_k v' \text{ and } v' \preceq_k v & \text{if and only if} & v \cong_k v' \\ v \preceq_k^{+1} v' \text{ and } v' \preceq_k^{+1} v & \text{if and only if} & v \cong_k^{+1} v'. \end{array}$$

Therefore, we concentrate on the first implication. As for $\mathrm{FO}^2(<)$, this an Ehrenfeucht-Fraïssé argument. We first define the Ehrenfeucht-Fraïssé game associated to $\Sigma_n(<)$ (*i.e.,* corresponding to $\preceq_k$) and then explain how to adapt it to $\preceq_k^{+1}$.

Ehrenfeucht-Fraïssé Game. The board of the $\Sigma_n(<)$-game consists of two words v, v' and there are two players called *Spoiler and Duplicator*. Moreover, initially, there exists a distinguished word among v, v' that we call the *active word* (this word may change as the game progresses). The game is set to last a predefined number k of rounds. When the game starts, both players have k pebbles. Finally, there is a parameter that gets updated during the game, a counter c called the *alternation counter*. Initially, c is set to 0. It may be incremented, but it has to remain bounded by $n-1$.

At the start of each round j, Spoiler chooses a word, either v or v'. Spoiler can always choose the active word, in which case both c and the active word remain unchanged. However, Spoiler can only choose the word that is not active when $c < n-1$, in which case the active word is switched and c is incremented by 1 (in particular, this may happen at most $n-1$ times). If Spoiler chooses v (resp. v'), he puts a pebble on a position x_j in v (resp. x'_j in v').

Duplicator must answer by putting a pebble at a position x'_j in v' (resp. x_j in v). Moreover, Duplicator must ensure that all pebbles that have been placed up to this point verify the following condition: for all $\ell_1, \ell_2 \leqslant j$, the labels at positions x_{ℓ_1}, x'_{ℓ_1} are the same, and $x_{\ell_1} < x_{\ell_2}$ if and only if $x'_{\ell_1} < x'_{\ell_2}$.

Duplicator wins if she manages to play for all k rounds, and Spoiler wins as soon as Duplicator is unable to play.

The $\Sigma_n(<, +1, min, max)$-game is defined similarly with the following additional constraint for Duplicator: at any time, for all ℓ_1, ℓ_2, $x_{\ell_1} = x_{\ell_2} + 1$ if and only if $x'_{\ell_1} = x'_{\ell_2} + 1$, $min(x_{\ell_1})$ if and only if $min(x'_{\ell_1})$ and $max(x_{\ell_1})$ if and only if $max(x'_{\ell_1})$.

▶ **Lemma 22** (Folklore). *For all $k \in \mathbb{N}$ and v, v', we have the following facts:*

- *$v \preceq_k v'$ iff Duplicator has a winning strategy in the k-round $\Sigma_n(<)$-game on v and v' with v as initial active word.*
- *$v \preceq_k^{+1} v'$ iff Duplicator has a winning strategy in the k-round $\Sigma_n(<, +1, min, max)$-game on v and v' with v as initial active word.*

In view of Lemma 22, we prove that $\lceil \mathfrak{u} \rceil_{2^{k+1}} \preceq_k^{+1} \lceil \mathfrak{u}' \rceil_{2^{k+1}}$ by giving a winning strategy for Duplicator in the corresponding k-round $\Sigma_n(<, +1, min, max)$-game. We call $\mathcal{G}$ this game. Duplicator's strategy involves playing a second shadow $\Sigma_n(<)$-game $\mathcal{S}$ on $\mathfrak{u}$ and $\mathfrak{u}'$, on which, by hypothesis and Lemma 20, she has a winning strategy in k rounds. We begin by setting up some notation that will help us define Duplicator's strategy.

Notation. Set $u = \lceil \mathfrak{u} \rceil_{2^{k+1}}$ and $u' = \lceil \mathfrak{u}' \rceil_{2^{k+1}}$. Assuming that $\mathfrak{u} \preceq_k \mathfrak{u}'$, we need to prove that $u \preceq_k^{+1} u'$. If $\mathfrak{u} \in S$ or $\mathfrak{u}' \in S$, then $\mathfrak{u} = \mathfrak{u}' = s \in S$ (again, the only well-formed word that contains the letter $s \in S$ is s) and the result is immediate.

Otherwise, by hypothesis, the words $\mathfrak{u}$ and $\mathfrak{u}'$ are of the form

$$\begin{aligned} \mathfrak{u} &= (s_0, e_1)(e_1, s_1, e_2) \cdots (e_m, s_m) \\ \mathfrak{u}' &= (s'_0, e'_1)(e'_1, s'_1, e'_2) \cdots (e'_{m'}, s'_{m'}) \end{aligned}$$

In particular, observe that since $\mathfrak{u} \preceq_k \mathfrak{u}'$ and the labels of the leftmost and rightmost positions are unique in $\mathfrak{u}$ and $\mathfrak{u}'$, we have $s_0 = s'_0$, $e_1 = e'_1$, $e_m = e'_{m'}$ and $s_m = s'_{m'}$. For the sake of simplifying the presentation, we assume that for all $i \leqslant m$, we have $\lceil s_i \rceil = a_i \in A$ and $\lceil e_i \rceil = b_i \in A$ (this does not harm the generality of the proof). Similarly, for all $i \leqslant m'$,

we assume that $\lceil s'_i \rceil = a'_i \in A$ and $\lceil e'_i \rceil = b'_i \in A$. By definition, we have

$$\begin{aligned} u &= \lceil \mathfrak{u} \rceil_{2^{k+1}} = a_0 (b_1)^{2^{k+1}} a_1 (b_2)^{2^{k+1}} \cdots (b_m)^{2^{k+1}} a_m \\ u' &= \lceil \mathfrak{u}' \rceil_{2^{k+1}} = a'_0 (b'_1)^{2^{k+1}} a'_1 (b'_2)^{2^{k+1}} \cdots (b'_{m'})^{2^{k+1}} a'_{m'}. \end{aligned}$$

Winning Strategy. We define an invariant $\mathcal{I}(\ell)$ (ℓ is the number of remaining rounds in the main game) that Duplicator has to satisfy when playing.

As she plays, Duplicator associates to each position $i \in \mathfrak{u}$, (resp. $i' \in \mathfrak{u}'$) a set of positions in u (resp. u') called the set of *marked positions* of i (resp. i'). All marked positions for i (resp. i') must belong to the b_i, a_i or b_{i+1} (resp. $b'_{i'}$, $a'_{i'}$ or $b_{i'+1}$) positions in u (resp. u'). Initially, all a_i (resp. $a'_{i'}$) are marked for i (resp. i'). Duplicator may define more positions as marked as the game progresses. All these new marked positions will be positions holding pebbles in $\mathcal{G}$.

Assume that there are ℓ rounds left to play and that pebbles have already been placed on u, u' in the main game $\mathcal{G}$ and on $\mathfrak{u}, \mathfrak{u}'$ in $\mathcal{S}$ in a way that satisfies the conditions of both Ehrenfeucht-Fraïssé games. We denote by $c_{\mathcal{G}}$ the alternation counter of the main game $\mathcal{G}$ and by $c_{\mathcal{S}}$ that of the shadow game $\mathcal{S}$. For all $i \in \mathfrak{u}$ (resp. $i' \in \mathfrak{u}'$) we denote by $x_1(i) < \cdots < x_{m_i}(i)$ (resp. $x'_1(i') < \cdots < x'_{m_{i'}}(i')$) the marked positions of i (resp. i'). Then $\mathcal{I}(\ell)$ holds if the following properties hold:

1. Duplicator has a winning strategy for playing at least ℓ more rounds in $\mathcal{S}$. Furthermore, either $c_{\mathcal{S}} > c_{\mathcal{G}}$, or $c_{\mathcal{S}} = c_{\mathcal{G}}$ and the active words in $\mathcal{S}$ and $\mathcal{G}$ are either $\mathfrak{u}$ and u, or $\mathfrak{u}'$ and u'.
2. Any position $x \in u$ ($x' \in u'$) that holds a pebble in $\mathcal{G}$ is marked for some $i \in \mathfrak{u}$ (resp. $i' \in \mathfrak{u}'$) holding a pebble in $\mathcal{S}$. Conversely, any position that is marked for $i \in \mathfrak{u}$, (resp. $i' \in \mathfrak{u}'$) is either a_i (resp. $a'_{i'}$) or a position holding a pebble in $\mathcal{G}$.
3. Let i, i' be positions of $\mathfrak{u}, \mathfrak{u}'$ on which there are corresponding pebbles in $\mathcal{S}$. Observe that since i, i' have the same label, $a_i = a'_{i'}$, $b_i = b'_{i'}$ and $b_{i+1} = b'_{i'+1}$.
   In that case, $m_i = m_{i'}$ and for all $j \leqslant m_i$, $x_j(i)$ is the $a_i = a'_{i'}$ position iff $x'_j(i')$ is the $a_i = a'_{i'}$ position and $x_j(i)$ holds a pebble of $\mathcal{G}$ iff $x'_j(i')$ holds the corresponding pebble. Finally, given $j < m_i$, let d and d' be the number of positions that are strictly between $x_j(i)$ and $x_{j+1}(i)$ (resp. $x'_j(i')$ and $x'_{j+1}(i')$). Note that by the condition above these positions are all labeled by $b_i = b'_{i'}$, or all labeled by $b_{i+1} = b'_{i'+1}$. We require that either $d = d'$, or $d \geqslant 2^\ell$ and $d' \geqslant 2^\ell$.
4. For all $i \in \mathfrak{u}$ (resp. $i' \in \mathfrak{u}'$), we have $x_{m_i}(i) < x_1(i+1)$ (resp. $x'_{m_{i'}}(i') < x'_1(i'+1)$). Moreover, there are more than $2^{\ell+1}$ copies of b_{i+1} (resp. $b'_{i'+1}$) that are strictly between these two positions.

It is clear that $\mathcal{I}(k)$ holds before the initial round. Assume now that there are $\ell + 1$ rounds left to play and that $\mathcal{I}(\ell + 1)$ holds. We explain how Duplicator can play in order to enforce $\mathcal{I}(\ell)$. Assume that Spoiler puts a pebble at a position $x \in u$ in $\mathcal{G}$ (the case when Spoiler plays in u' is symmetric). We distinguish two cases depending on the position x.

There exists $i \in \mathfrak{u}$ such that $x_1(i) \leqslant x \leqslant x_{m_i}(i)$. If there is already a pebble on i in $\mathcal{S}$, then we set i' as the position holding the corresponding pebble in $\mathfrak{u}'$. Otherwise, Duplicator simulates a move by Spoiler in $\mathcal{S}$ by putting a pebble on position i. We set i' as the answer she obtains from her strategy in $\mathcal{S}$. Note that by hypothesis all pebbles in $\mathfrak{u}, \mathfrak{u}'$ (including i, i') satisfy the conditions of the $\Sigma_n(<)$-game.

If x is already a marked position $x_j(i)$ of i, then Duplicator answer by putting a corresponding pebble on $x'_j(i')$. Note that this answer is correct by hypothesis on i, i' for the

$\Sigma_n(<)$-game and by hypothesis on the marked positions of i, i' as stated in Item 2 of $\mathcal{I}(\ell+1)$. Since both positions were already marked for i, i', it is then simple to verify that $\mathcal{I}(\ell)$ holds.

Assume now that x is not yet marked. Then x is a b_i or a b_{i+1} position (a_i positions are aways marked). Assume that x is a b_i position (the case b_{i+1} is similar). Recall that $m_i = m'_{i'}$ by Item 2 in $\mathcal{I}(\ell+1)$. Let j such that $x_j(i) < x < x_{j+1}(i)$. By Item 3 of $\mathcal{I}(\ell+1)$ it is immediate than one can find an answer $x' \in u'$ such that $x'_j(i') < x' < x'_{j+1}(i')$ and Item 3 of $\mathcal{I}(\ell)$ remains satisfied with x, x' as new marked positions of i, i'. Again this answer is correct by hypothesis on i, i' for the $\Sigma_n(<)$-game and by hypothesis on the marked positions of i, i' as stated in Item 2 of $\mathcal{I}(\ell+1)$. It is then simple to verify that $\mathcal{I}(\ell)$ remains satisfied.

There exists $i \in \mathbb{u}$ such that $x_{m_{i-1}}(i-1) < x < x_1(i)$. From Item 4 in $\mathcal{I}(\ell+1)$, we know that there are at least $2^{\ell+2}$ copies of b_i between $x_{m_{i-1}}(i-1)$ and $x_1(i)$. It follows, that there are either at least $2^{\ell+1}$ copies of b_i between $x_{m_{i-1}}(i-1)$ and x or at least $2^{\ell+1}$ copies of b_i between x and $x_1(i)$. Since both cases are symmetric, assume that we are in the first case: there are at least $2^{\ell+1}$ copies of b_i between $x_{m_{i-1}}(i-1)$ and x.

If there is already a pebble on i in $\mathcal{S}$, then we set i' as the position holding the corresponding pebble in $\mathbb{u}'$. Otherwise, Duplicator simulates a move by Spoiler in $\mathcal{S}$ by putting a pebble on position i. We set i' as the answer she obtains from her strategy in $\mathcal{S}$. Note that by hypothesis all pebbles in $\mathbb{u}, \mathbb{u}'$ (including i, i') satisfy the condition of the $\Sigma_n(<)$-game.

Set d as the number of copies of b_i between x and $x_1(i)$, *i.e.*, $x = x_1(i) - (d+1)$. If $d < 2^\ell$, we set $x' \in u'$ as the position $x' = x_1(i') - (d+1)$. Otherwise we set $x' \in u'$ as the position $x' = x_1(i) - (2^\ell + 1)$. In both cases, x' is Duplicator's answer and we set x, x' as new marked positions of i, i'. Note that this answer is correct by hypothesis on i, i' for the $\Sigma_n(<)$-game. It is immediate that Item 1 is satisfied in $\mathcal{I}(\ell)$. Item 2, 3 and 4 of $\mathcal{I}(\ell)$ are satisfied by choice of x'.

C Proof of Theorem 15

This section is divided in three parts. In the first one, we recall the formal definition of the semidirect product operation. In the next two ones, we prove both directions of Theorem 15.

C.1 The semidirect product

Let M be an ordered monoid and T be an ordered semigroup. A *semidirect product* of M and T is an operation which is parametrized by an *action* of T on M and outputs a new ordered semigroup, whose base set is $M \times T$. In particular, one can obtain different semidirect products out of the same M and T, depending on the chosen action.

Let '$+$' and '$\cdot$' be the operations of M and T respectively. Note that we choose to denote the operation on M additively. This is for the sake of simplifying the presentation. However, this does not mean that we assume M to be commutative. An action '$\cdot$' of T on M is a mapping $(t, s) \mapsto t \cdot s$ from $T^1 \times M$ to M such that, for all $s, s' \in M$ and all $t, t' \in T$:

1. $t \cdot (t' \cdot s) = (t \cdot t') \cdot s$.
2. $t \cdot (s + s') = t \cdot s + t \cdot s'$.
3. $1_T \cdot s = s$.
4. if $s \leqslant s'$, then $t \cdot s \leqslant t \cdot s'$.
5. if $t \leqslant t'$, then $t \cdot s \leqslant t' \cdot s$.
6. $t \cdot 1_M = 1_M$.

Given a fixed action '$\cdot$' of T on M, the *semidirect product $M * T$ of M and T with respect*

to the action '$\cdot$' is the set $M \times T$ equipped with the following operation:

$$(s,t) \cdot (s',t') = (s + t \cdot s', t \cdot t')$$

and the product order: $(s,t) \leqslant (s',t')$ if $s \leqslant s'$ and $t \leqslant t'$. One can verify that this does yield an ordered semigroup [15].

We will only use the semidirect product with semigroups $T \in \mathsf{D}$. In particular, if V is a variety of ordered monoids, we denote by $\mathsf{V} * \mathsf{D}$ the variety of ordered semigroups generated by all semidirect products $M * T$, for some $M \in \mathsf{V}$ and some $T \in \mathsf{D}$.

We are now ready to prove Theorem 15. Recall that we have a non-trivial variety V of ordered monoids, two languages L and L' recognized by a morphism $\alpha : A^+ \to S$, and $\mathbb{L}, \mathbb{L}' \subseteq \mathbb{A}_\alpha^+$ the associated languages of well-formed words.

We prove that L is $(\mathsf{V} * \mathsf{D})$-separable from L' if and only if $\mathbb{L}$ is V-separable from $\mathbb{L}'$. We prove each direction in its own subsection.

C.2 From $(\mathsf{V} * \mathsf{D})$-separability to V-separability

We prove that if L is $(\mathsf{V} * \mathsf{D})$-separable from L', then $\mathbb{L}$ is V-separable from $\mathbb{L}'$. Note that we reuse the construction which associates a canonical word $\lceil \mathbb{w} \rceil_i \in A^+$ to every word $\mathbb{w} \in \mathbb{A}_\alpha^+$ and natural $i \geqslant 1$ (see Section 4.1 for details).

Assume that L is $(\mathsf{V} * \mathsf{D})$-separable from L'. This means that there exists an element of $(\mathsf{V} * \mathsf{D})$ separating L and L'. By [15, Prop. 3.5], such an ordered semigroup is an ordered quotient of an ordered subsemigroup of a semidirect product $M * T$, with $M \in \mathsf{V}$ and $T \in \mathsf{D}$. Therefore, $M * T$ itself separates L and L'. Hence, there is some upward closed $F \subseteq M * T$ and a morphism $\delta : A^+ \to M * T$ such that $\delta^{-1}(F)$ separates L from L'.

We construct a separator in V for $\mathbb{L}$ and $\mathbb{L}'$. Set $T = \{t_1, \dots, t_n\}$ and observe that since V is non-trivial, it contains an ordered monoid N containing at least n distinct elements. We choose n such elements $t'_1, \dots, t'_n$ of N. The choice is essentially arbitrary, but we ask $t'_1, \dots, t'_n$ to be pairwise incomparable with respect to the partial order $\leqslant$. We prove that $\mathbb{L}$ can be separated from $\mathbb{L}'$ using the ordered monoid $\mathbb{M} = M \times N \in \mathsf{V}$ (recall that a variety is closed under cartesian product). For an element $t = t_i$ of T, we denote by t' the element t'_i of N.

We define a morphism $\gamma : \mathbb{A}_\alpha^+ \to \mathbb{M}$ as follows. Let ω be the idempotent power $\omega(M * T)$ of $M * T$. Set $\mathbb{a} = (e, s, f) \in \mathbb{A}_\alpha$, so that $\lceil \mathbb{a} \rceil_\omega = w_e^\omega w_s w_f^\omega$. Let $\delta(w_e^\omega) = (m_e, t_e) \in M * T$ and $\delta(\lceil \mathbb{a} \rceil_\omega) = (m, t)$. We define $\gamma(\mathbb{a}) \in M \times N$ as follows:

$$\gamma(\mathbb{a}) = \begin{cases} (t_e \cdot m,\ t') & \text{when } f = 1_S, \\ (t_e \cdot m,\ 1_N) & \text{otherwise.} \end{cases}$$

This defines a morphism $\gamma : \mathbb{A}_\alpha \to M \times N \in \mathsf{V}$. It remains to prove that γ recognizes a separator of $\mathbb{L}$ and $\mathbb{L}'$. This is a consequence of the next lemma.

▶ **Lemma 23.** *Let $\mathbb{w} \in \mathbb{A}_\alpha^+$ be well-formed, and set $(m, t_i) = \delta(\lceil \mathbb{w} \rceil_\omega)$. Then $\gamma(\mathbb{w}) = (m, t'_i)$.*

Before proving the lemma, we use it to conclude the proof. Define $\mathbb{F} \subseteq \mathbb{M}$ by $\mathbb{F} = \{(m, t'_i) \mid (m, t_i) \in F\}\}$. One can verify that $\mathbb{F}$ is upward closed. We claim that $\gamma^{-1}(\mathbb{F})$ separates $\mathbb{L}$ from $\mathbb{L}'$.

Assume first that $\mathbb{w} \in \mathbb{L}$. By Fact 8, $\lceil \mathbb{w} \rceil_\omega \in L$, hence $\delta(\lceil \mathbb{w} \rceil_\omega) \in F$. It then follows from Lemma 23 that $\gamma(\mathbb{w}) \in \mathbb{F}$. Conversely if $\mathbb{w} \in \mathbb{L}'$, we have $\delta(\lceil \mathbb{w} \rceil_\omega) \notin F$. It then follows from Lemma 23 that $\gamma(\mathbb{w}) \notin \mathbb{F}$ which terminates the proof. We now prove Lemma 23.

Proof of Lemma 23. We first show that the first component in M of $\delta(\lceil \mathbb{w} \rceil_\omega)$ and of $\gamma(\mathbb{w})$ are equal. The proof consists in a straightforward but tedious computation. Set $\mathbb{w} = \mathbb{a}_1 \cdots \mathbb{a}_p \in \mathbb{A}_\alpha^+$ that is well-formed. Set $\mathbb{a}_i = (e_{i-1}, s_i, e_i)$ and recall that, in view of the definition of $\lceil \mathbb{w} \rceil_i$ given in Section 4.1, we have chosen words w_{e_i} and w_{s_i} such that:

$$\lceil \mathbb{a}_i \rceil_\omega = w_{e_{i-1}}^\omega \cdot w_{s_i} \cdot w_{e_i}^\omega$$

For each idempotent $e = e_i$, set $\delta(w_e^\omega) = (m_e, t_e) \in M * T$ and for each element $s = s_i$, let $\delta(w_s) = (m_s, t_s) \in M * T$. Note that by definition of ω, the element $(m_e, t_e) = \delta(w_e^\omega) = \delta(w_e)^\omega$ is idempotent, so $(m_e, t_e) = (m_e + t_e \cdot m_e, t_e^2)$. In particular, t_e is idempotent in T. Further, we have for all i:

$$m_{e_i} + t_{e_i} \cdot m_{e_i} = m_{e_i}. \tag{3}$$

To lighten the notation, from now on, let us write the action of T on M as tm instead of $t \cdot m$. For each $\mathbb{a}_i = (e_{i-1}, s_i, e_i)$, we then have

$$\begin{aligned}
\delta(\lceil \mathbb{a}_i \rceil_\omega) &= \delta(w_{e_{i-1}}^\omega)\delta(w_{s_i})\delta(w_{e_i})^\omega \\
&= (m_{e_{i-1}}, t_{e_{i-1}})(m_{s_i}, t_{s_i})(m_{e_i}, t_{e_i}) \\
&= \Big(m_{e_{i-1}} + t_{e_{i-1}} m_{s_i} + t_{e_{i-1}} t_{s_i} m_{e_i}, \quad t_{e_i}\Big)
\end{aligned} \tag{4}$$

where, for computing the 2nd component, we used the fact that t_{e_i} is idempotent in $T \in \mathsf{D}$. Similarly, by definition we have $\lceil \mathbb{w} \rceil_\omega = (w_{e_0})^\omega w_{s_1} (w_{e_1})^\omega \cdots (w_{e_{p-1}})^\omega w_{s_p} (w_{e_p})^\omega$, and

$$\begin{aligned}
\delta(\lceil \mathbb{w} \rceil_\omega) &= \delta(w_{e_0}^\omega)\delta(w_{s_1})\delta(w_{e_1})^\omega w \cdots \delta(w_{e_{p-1}})^\omega \delta(w_{s_p}) w_{s_p} (w_{e_p})^\omega \\
&= (m_{e_0}, t_{e_0})(m_{s_1}, t_{s_1})(m_{e_1}, t_{e_1}) \cdots (m_{e_{p-1}}, t_{e_{p-1}})(m_{s_p}, t_{s_p})(m_{e_p}, t_{e_p}) \\
&= \Big(m_{e_0} + (t_{e_0} m_{s_1} + t_{e_0} t_{s_1} m_{e_1}) + \cdots + (t_{e_{p-1}} m_{s_p} + t_{e_{p-1}} t_{s_p} m_{e_p}), \quad t_{e_p}\Big).
\end{aligned} \tag{5}$$

Again, for the last equality, we used the definition of the semidirect product and the fact that each t_{e_i} is an idempotent in T, which implies, since $T \in \mathsf{D}$, that $t \cdot t_{e_i} = t_{e_i}$ for all $t \in T$.

Using (3) for each i, one can replace m_{e_i} in (5) by $m_{e_i} + t_{e_i} m_{e_i}$. Taking into account that t_{e_i} is idempotent in T, this yields for this first component of $\delta(\lceil \mathbb{w} \rceil_\omega)$ the value

$$m_{e_0} + t_{e_0} m_{e_0} + (t_{e_0} m_{s_1} + t_{e_0} t_{s_1} m_{e_1} + t_{e_1} m_{e_1}) + \cdots + (t_{e_{p-1}} m_{s_p} + t_{e_{p-1}} t_{s_p} m_{e_p} + t_{e_p} m_{e_p})$$

Observe that since $\mathbb{w}$ is well-formed, $e_0 = 1_S$, hence $m_{e_0} = 1_M$, which is the neutral element for the '+' operation on M. In the same way, $e_p = 1_S$, hence $m_{e_p} = 1_M$, and therefore, using the last axiom of an action, we deduce that $t_{e_p} m_{e_p} = 1_M$. Hence, these two elements can be removed from the expression of the first component of $\delta(\lceil \mathbb{w} \rceil_\omega)$. Therefore, this first component can be rewritten, using associativity, as:

$$(t_{e_0} m_{e_0} + t_{e_0} m_{s_1} + t_{e_0} t_{s_1} m_{e_1}) + \cdots + (t_{e_{p-1}} m_{e_{p-1}} + t_{e_{p-1}} m_{s_p} + t_{e_{p-1}} t_{s_p} m_{e_p}). \tag{6}$$

On the other hand, in view of (4) and by definition of γ, the first component of $\gamma(\mathbb{a}_i) \in M \times N$ is

$$(t_{e_{i-1}} \cdot (m_{e_{i-1}} + t_{e_{i-1}} m_{s_i} + t_{e_{i-1}} t_{s_i} m_{e_i}) = (t_{e_{i-1}} m_{e_{i-1}} + t_{e_{i-1}} m_{s_i} + t_{e_{i-1}} t_{s_i} m_{e_i}). \tag{7}$$

Therefore, one can compute the first component of $\gamma(\mathbb{w}) = \gamma(\mathbb{a}_1 \cdots \mathbb{a}_p) = \gamma(\mathbb{a}_1) \cdots \gamma(\mathbb{a}_p)$ by summing the values (7) for $i = 1, \ldots, p$ (recall that the operation on M is noted additively),

which gives the value computed in (6). Hence we have shown that the first component in M of $\delta(\lceil \mathbb{w} \rceil_\omega)$ and of $\gamma(\mathbb{w})$ are equal.

It remains to check that when the second component of $\delta(\lceil \mathbb{w} \rceil_\omega)$ is equal to some $t \in T$, then the second component of $\gamma(\mathbb{w})$ is the corresponding element $t' \in N$. This is simpler: by definition of a well-formed word, we have $e_i \neq 1_S$ for $i < p$, and $e_p = 1_S$. By definition of γ, it follows that the second component of $\gamma(\mathbb{w})$ is the second component of $\gamma(\mathbb{a}_p)$, namely t'_p. Now, since $T \in \mathsf{D}$, the second component of $\delta(\lceil \mathbb{a} \rceil_\omega)$ is t_p, which concludes the proof. ◀

C.3 From V-separability to $(\mathsf{V} * \mathsf{D})$-separability

We prove that if $\mathbb{L}$ is V-separable from $\mathbb{L}'$, then L is $(\mathsf{V} * \mathsf{D})$-separable from L'. Note that we reuse the construction which associates to every word $w \in A^+$ a canonical word $\lfloor w \rfloor \in \mathbb{A}_\alpha^+$ (see Section 4.2 for details).

Assume that $\mathbb{L}$ is V-separable from $\mathbb{L}'$. This means that we have a morphism $\gamma : \mathbb{A}_\alpha^* \to \mathbb{M}$ with $\mathbb{M}$ an ordered monoid in V and $\mathbb{F} \subseteq \mathbb{M}$ upward-closed such that $\gamma^{-1}(\mathbb{F})$ separates $\mathbb{L}$ from $\mathbb{L}'$. We need to construct a separator in $\mathsf{V} * \mathsf{D}$ for L and L'. The main idea is to define a morphism, which given $w \in A^+$, computes $\gamma(\lfloor w \rfloor)$. This is slightly technical however as the morphism needs some machinery to make this computation.

We begin with some notations. To every word $w \in A^+$, we associate an element $lab(w) \in \mathbb{M}$. Let x be the last position in w and consider the construction of $\lfloor w \rfloor$. If x is distinguished, we set $lab(w) = \gamma(\mathbb{a})$ with $\mathbb{a}$ the label of $\lfloor x \rfloor$ in $\lfloor w \rfloor$. Otherwise, we simply set $lab(w) = 1_{\mathbb{M}}$. We can now start the construction of our separator. We have to define the following objects:

- An ordered semigroup $T \in \mathsf{D}$.
- An ordered monoid $M \in \mathsf{V}$.
- An action of T on M yielding a semidirect product $M * T$.
- A morphism $\delta : A^+ \to M * T$ which recognizes the desired separator.

Definition of T. We set T as the set $\{w \in A^+ \mid |w| \leqslant 2|S|\}$ equipped with the following operation. If $w, w' \in T$, we set $w \cdot w'$ as the suffix of length $2|S|$ of the word ww' when ww' has length $\geqslant 2|S|$ and as ww' otherwise. One can verify that this operation is indeed associative and that $T \in \mathsf{D}$. We use equality as the partial order on T.

Observe that we have a natural morphism $\rho : A^+ \to T$ such that $\rho(w)$ is w if $|w| \leqslant 2|S|$, and $\rho(w)$ is the suffix of length $2|S|$ of w otherwise. Observe that by Lemma 11, we have the following fact.

▶ **Fact 24.** *For every $w \in A^+$, $lab(w) = lab(\rho(w))$.*

Definition of M. We set $M \in \mathsf{V}$ as the cartesian product $\mathbb{M}^{T^1}$ (recall that as a variety of ordered monoids, V is closed under cartesian product).

▶ Remark. Since we intend to take a semidirect product of M and T, we will denote the semigroup operations of both M and $\mathbb{M}$ additively in order to clarify the presentation.

Definition of $M * T$. If $w \in T$ and $f \in M$ (*i.e.*, f is a mapping $f : T^1 \to \mathbb{M}$), we set $w \cdot f$ as the mapping $g : T^1 \to M$ such that $g(u) = f(u \cdot w)$. One can verify that '$\cdot$' is an action of T on M. In the remainder of the proof, we denote by $M * T$ the semidirect product of M and T with respect to this action.

Definition of δ. Set $f_{Id} : T^1 \to \mathbb{M}$ defined as follows. We set $f_{Id}(1_T) = 1_{\mathbb{M}}$ and $f_{Id}(w) = lab(w)$ when $w \in T$. We can now define $\delta : A^+ \to M * T$. Let $a \in A^+$, we set $\delta(a)$ as the

pair (f_a, a) where $f_a = a \cdot f_{Id}$, *i.e.*, the mapping $f_a : w \mapsto f_{Id}(wa)$. It now remains to prove that δ does recognize a separator of L from L'. This is a consequence of the following lemma.

▶ **Lemma 25.** *Let $w \in A^+$, $(f, u) = \delta(w)$ and $end(u)$ as the label of the last position in $\lfloor u \rfloor$. Then,*

$$\gamma(\lfloor w \rfloor) = f(1_T) \cdot \gamma(end(u)).$$

We first use the lemma to conclude the proof. Set $F \subseteq M * T$ as the set

$$F = \{(f, u) \mid f(1_T) \cdot \gamma(end(u)) \in \mathbb{F}\}.$$

One can verify that F is upward closed (this is essentially because $\mathbb{F}$ is upward-closed). It is immediate from Lemma 25 that $\delta(w) \in F$ iff $\gamma(\lfloor w \rfloor) \in \mathbb{F}$. We claim that $\delta^{-1}(F)$ separates L from L'.

Assume first that $w \in L$, we need to prove that $w \in \delta^{-1}(F)$. By Fact 10, we have $\lfloor w \rfloor \in \mathbb{L}$, hence $\gamma(\lfloor w \rfloor) \in \mathbb{F}$ and $\delta(w) \in F$. Similarly, if $w \in L'$, $\lfloor w \rfloor \in \mathbb{L}'$, hence $\gamma(\lfloor w \rfloor) \notin \mathbb{F}$ and $\delta(w) \notin F$ which terminates the proof. It finally remains to prove Lemma 25.

Proof of Lemma 25. Set $w = a_1 \cdots a_n$ and $\lfloor w \rfloor = \mathbb{a}_1 \cdots \mathbb{a}_m$. By definition, we have:

$$f = \rho(a_1) \cdot f_{Id} + \rho(a_1 a_2) \cdot f_{Id} + \cdots + \rho(a_1 \cdots a_n) \cdot f_{Id}$$

By definition of f_{Id} and by Fact 24 this means that:

$$f(1_T) = lab(a_1) + lab(a_1 a_2) + \cdots + lab(a_1 \cdots a_n)$$

It is then immediate from the definition of $\lfloor w \rfloor$ that $f(1_T) = \gamma(\mathbb{a}_1 \cdots \mathbb{a}_{m-1})$. Hence $\gamma(\lfloor w \rfloor) = f(1_T) \cdot \gamma(end(w))$. This finishes the proof since u is the suffix of length $2|S|$ of w, and therefore $end(u) = end(w)$ by Lemma 11. ◀