Adding successor: A transfer theorem for separation and covering

THOMAS PLACE, LaBRI, Bordeaux University

MARC ZEITOUN, LaBRI, Bordeaux University

Given a class C of word languages, the C-separation problem asks for an algorithm that, given as input two regular languages, decides whether there exists a third language in C containing the first language, while being disjoint from the second. Separation is usually investigated as a means to obtain a deep understanding of the class C.

In the paper, we are mainly interested in classes defined by logical formalisms. Such classes are often built on top of each other: given some logic, one builds a stronger one by adding new predicates to its signature. A natural construction is to enrich a logic with the successor relation. In this paper, we present a transfer result applying to this construction: we show that for suitable logically defined classes, separation for the logic enriched with the successor relation reduces to separation for the original logic. Our theorem also applies to a problem that is stronger than separation: covering. Moreover, we actually present two reductions: one for languages of finite words and the other for languages of infinite words.

1 INTRODUCTION

Context. A central problem in formal languages theory 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 [5, 6, 11, 37, 43, 45]. In this paper, we investigate fragments of MSO over finite and infinite words. In this context, understanding the expressive power of a fragment is often associated to a decision problem: *membership*. Given a logical fragment, one may associate the class C of all word languages that can be defined by a sentence of this fragment. When C is such a class, the C-membership problem asks for a decision procedure that tests whether some input regular language belongs to C. Intuitively, setting such an algorithm requires a deep understanding of C: it involves considering *all* languages within C.

Membership has been solved for many natural fragments of MSO, the most prominent one being FO(<): first-order logic equipped with a predicate "<" for the linear ordering. For finite words, the solution was found by Schützenberger, McNaughton and Papert [18, 38]. They characterized the regular languages that are definable in FO(<) by a syntactic, easily decidable property on a canonical recognizer of this language (such as its minimal automaton or its syntactic monoid). This result was later generalized to infinite words by Perrin [19]. It now serves as a commonly followed template, which was used successfully to solve membership for many other logical formalisms.

Research on this topic is still ongoing and membership remains open for several fragments. A prominent example is the quantifier alternation hierarchy of first-order logic, which classifies it into

levels $\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<)$. Despite years of investigation, only the lower levels have been solved by Simon [39], Pin and Weil [22, 24], the authors [29] and the first author [26, 27]. Furthermore, making progress has often required moving beyond the standard approach to membership questions. The latest results for the levels $\mathcal{B}\Sigma_2(<)$, $\Sigma_3(<)$ and $\Sigma_4(<)$ are based on decision problems that are stronger than membership: *separation* and *covering*. Given a class C, the C-*separation problem* asks for a decision procedure that takes *two* input regular languages and tests whether there exists a third one in C containing the first language while being disjoint from the second one. Covering, which we defined in [32, 34] is even more general: it takes two different objects as input: a regular language L and a finite set of regular languages $\mathbf{L}$. It asks whether there exists a C-cover $\mathbf{K}$ of L (*i.e.*, a finite set of languages in C whose union includes L) such that no language $K \in \mathbf{K}$ intersects all languages in $\mathbf{L}$. Separation is just the special case when $\mathbf{L}$ is a singleton. Both problems are decidable for FO($<$) as we showed [30, 33], both for finite and infinite words.

Because of these results, separation and covering have quickly replaced membership as the central question when trying to "understand" a given class of languages. However, the main motivation for considering separation and covering is more profound: while harder than membership, they are also more rewarding with respect to the knowledge gained on the investigated class C. Intuitively, a membership algorithm only yields benefits for the languages of C: we are able to detect them and to build a description witnessing this membership. On the other hand, separation and covering algorithms are universal: their benefits apply to **all** languages. An insightful point of view is to see them as *approximation problems*. For example, given an input pair (L_1, L_2), the objective of separation is to over-approximate L_1 by a language in C while L_2 is the specification of what an acceptable approximation is.

In the paper, we investigate separation and covering for several natural fragments of FO($<$). Specifically, we consider the levels $\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<)$ in the quantifier alternation hierarchy and the two-variable fragment FO$^2(<)$. However, we shall not work with these fragments themselves. Instead, we are interested in stronger variants which are built from them in a natural way. A crucial observation is that for these fragments, the drop in expressive power forbids the use of natural relations that could be defined from the linear order in full first-order logic. The main example is "+1": the *successor relation*. While FO($<$) is powerful enough to express it ("$x + 1 = y$" is equivalent to "$x < y \wedge \neg\exists z(x < z < y)$"), this is not the case for FO$^2(<)$, $\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<)$. Hence, there are two natural variants for each of these fragments: a weak one which is only equipped with the linear ordering (denoted FO$^2(<)$, $\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<)$) and a strong one which is equipped with additional predicates such as successor (denoted FO$^2(<, +1)$, $\Sigma_n(<, +1)$ and $\mathcal{B}\Sigma_n(<, +1)$). Our objective in this paper is to investigate separation and covering problems associated to strong variants.

State of the art. Naturally, these strong logical fragments were first investigated using the membership problem. However, this proved to be unexpectedly difficult. In most cases, even when the weak variant is known to have decidable membership, proving that this is also the case for the strong one can be highly nontrivial. Examples include the membership proofs of $\mathcal{B}\Sigma_1(<, +1)$ and $\Sigma_2(<, +1)$, which involve difficult and intricate combinatorial arguments [13, 14, 16] or a wealth of algebraic machinery [24, 25]. Another issue is that most proofs directly deal with the strong variant. 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 would be to first obtain a solution of the decision problems for the weak variant before lifting it to the strong one via a *generic transfer result*.

This idea has first been investigated by Straubing [42] for the membership problem in the setting of finite words. He chose to formulate his approach using algebraic terminology. It is known that any class of languages satisfying appropriate properties is characterized by some algebraic

variety V: a language is in the class if and only if its syntactic monoid belongs to V. This result is the variety theorem of Eilenberg [10]. Straubing's approach was to capture the intuitive connection between weak and strong fragments using a generic operation on algebraic varieties called *wreath product*. Though this is nontrivial, it has been shown that for most logical fragments (including the ones we consider in the paper), if V is the variety corresponding to the weak variant, then the strong one corresponds to the variety V ∘ D: the wreath product of V with D (where the D is a fixed variety). Thus, Straubing's approach was to show that the operation V ↦ V ∘ D preserves the decidability of membership.

Unfortunately, this is not true in general [2]. In fact, while decidability is preserved for all natural 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. It turns out that in the separation setting, this approach is more robust: it has been shown by Steinberg [41] that decidability of separation is preserved by the operation V ↦ V ∘ D. However, this result has several downsides:

- Steinberg's theorem is not about separation: it states a purely algebraic property of varieties of the form V ∘ D (they have "decidable pointlikes"). The connection with separation is indirect and made with another result by Almeida [1]. Therefore, while interesting when already starting from algebra, this approach is less satisfying from a logical point of view: it 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. One has to manipulate three objects of different nature simultaneously: logic, classes of languages and algebraic varieties. Proofs are also often nontrivial and require a deep understanding of complex objects, which may be scattered in the bibliography.

- Steinberg's result only applies to classes of languages closed under complement (which excludes the fragments Σ_n in the quantifier alternation hierarchy). This limitation is tied to the connection with algebraic varieties which only holds for classes closed under complement. While this connection may be lifted to a more general setting [23, 25], this requires introducing even more algebraic vocabulary.

- These results are specific to finite words while we intend to investigate both finite and infinite words.

Contributions. We present a new transfer theorem applying to all fragments presented above. For each of them, we show that separation and covering for the strong variant reduce to the same problem for the weak one. Our approach is generic and similar to the original one of Straubing described formerly. However, rather than choosing algebra to formulate it, we use a pure language theoretic point of view. Specifically, we define a product between classes of languages, called *enrichment*. Given two classes C and $\mathcal{D}$, it builds a new one denoted by $C \circ \mathcal{D}$: the $\mathcal{D}$-enrichment of C. As the notation suggests, this operation is designed as the language theoretic counterpart of the wreath product. We then show the two following properties:

(1) For all fragments that we consider, if C is the class corresponding to the weak variant, then the strong one corresponds $C \circ$ SU (SU is a fixed class: the *suffix languages*).

(2) Given any class C satisfying standard closure properties, covering and separation for $C \circ$ SU reduce to the same problem for C.

Using such a language theoretic approach has several important benefits over the algebraic one. Let us summarize them.

- The definition of enrichment is simple, and requires much less machinery than the wreath product. We avoid a lot of algebraic vocabulary, which we do not need. We only work with

two objects: logic and classes of languages. The only needed piece of algebra is the elementary definition of regular languages in terms of finite monoids.

- Our proof is self-contained and much simpler than previous ones. It only relies on basic notions on regular languages. A consequence is that our techniques yield much more intuition on the logical point of view.

- Enrichment makes sense for *any* class of language, even if it is not closed under complement. Furthermore, closure under complement is not required for applying our reduction theorem. Thus, contrary to [41] our results capture the Σ_n levels in the quantifier alternation hierarchy of first-order logic.

- Our definitions and proofs adapt smoothly to the setting of infinite words. We have two definitions of enrichment and two reduction theorems: the first are for classes of languages of finite words and the second for classes of languages of infinite words.

- In both settings of finite and infinite words, our results apply to two different problems: separation and covering.

It is already known that covering and separation are decidable for the weak variants of many logical fragments. Thus, when combining these algorithms with our results, we shall obtain new separation and covering procedures for several strong variants. Over words, it is known that both problems are decidable for $FO^2(<)$ [28, 32, 34], $\Sigma_1(<)$ [9, 32, 34], $\mathcal{B}\Sigma_1(<)$ [9, 28, 36], $\Sigma_2(<)$ [29, 35, 36], $\mathcal{B}\Sigma_2(<)$ [36] and $\Sigma_3(<)$ [26, 27]. Thus, we obtain the decidability of separation and covering for $FO^2(<, +1)$, $\Sigma_1(<, +1)$, $\mathcal{B}\Sigma_1(<, +1)$, $\Sigma_2(<, +1)$, $\mathcal{B}\Sigma_2(<, +1)$ and $\Sigma_3(<, +1)$ over words. Over infinite words the situation is more complicated: while the state of the art is roughly the same as for finite words, many of these results are yet unpublished. It was shown in [21] that separation is decidable for $\Sigma_2(<)$ and $\Sigma_3(<)$. Thus, we get that separation is decidable for $\Sigma_2(<, +1)$ and $\Sigma_3(<, +1)$ over infinite words.

Organization of the Paper. In Section 2, we set up the notation and present the separation and covering problems. In Section 3, we define the logical fragments that we investigate in the paper. Section 4 is devoted to our main theorem for languages of finite words: we define the enrichment operation on classes of languages of finite words, and we show that covering and separation for $C \circ SU$ reduce to the corresponding problem for C. The next two sections are devoted to applying this result to our logical fragments: we do so for two-variable first-order logic in Section 5 and the quantifier alternation hierarchy of first-order logic in Section 6. Finally, in Section 7, we generalize our results to the setting of infinite words: we adapt SU-enrichment for classes of languages of infinite words and we lift our reduction theorem to this setting.

This paper is the full version of [31]. From the conference version, the point of view has been changed from a purely logical one to a language theoretic one with the SU-enrichment operation. In particular, this means that while the underlying ideas are the same, the theorem presented in this full version is more general and applies to all classes built using SU-enrichment. Additionally, the reduction for infinite words is new.

2 PRELIMINARIES

In this section, we introduce the objects that we investigate in the paper. We first recall basic definitions about (finite and infinite) words and regular languages. Then, we present the two decision problems that we consider: covering and separation.

2.1 Words and classes of languages

An alphabet is a *finite* set A of symbols, which are called *letters*. We shall consider both finite and infinite words. Given some alphabet A, we denote by A^+ the set of all nonempty finite words and by A^* the set of all finite words over A (*i.e.*, $A^* = A^+ \cup \{\varepsilon\}$). Moreover, we write A^ω for the set of all infinite words over A. Note that we shall always use the term "word" to mean a finite word (*i.e.*, an element of A^*). On the other hand, we shall speak of an "ω-word" when considering an infinite word, *i.e.*, an element of A^ω. Finally, we let $A^\infty = A^* \cup A^\omega$.

If $u \in A^*$ and $v \in A^\infty$ we write $u \cdot v \in A^\infty$ or $uv \in A^\infty$ for the concatenation of u and v. Note that if $v \in A^*$, then $uv \in A^*$ and if $v \in A^\omega$, then $uv \in A^\omega$. We shall also consider infinite products. Let $(u_n)_{n \in \mathbb{N}}$ by a infinite family of words (*i.e.*, $u_n \in A^*$ for all $n \in \mathbb{N}$), then we may construct a new word or ω-word $u_0u_1u_2u_3\cdots \in A^\infty$ by concatenating them all. Observe that $u_0u_1u_2u_3\cdots \in A^\omega$ when there are infinitely many indices $n \in \mathbb{N}$ such that $u_n \neq \varepsilon$. Otherwise, $u_0u_1u_2u_3\cdots \in A^*$. Finally, when $u \in A^*$ is a single word, we denote by $u^\omega \in A^\infty$ the infinite concatenation $uuuu\cdots$.

The *length* of a word $u \in A^*$, denoted by $|u|$, is its number of letters. When $u \in A^\omega$ is an ω-word, we let $|u| = \infty$. Since we consider logic, we shall often view words and ω-words as linearly ordered sets of labeled *positions*: the domain of a word $u \in A^*$ is $\{0, \ldots, |u| - 1\}$, while the domain of an ω-word is simply $\mathbb{N}$. In particular, we shall use the following notation. Let u be a word or an ω-word and let i, j be two integers. We let $u[i, j] \in A^*$ be the following word:

(1) If $i \leq j \leq |u| - 1$, then $u[i, j]$ is the infix of u obtained by keeping all positions from i to j in u. For example, if $u = a_0 \cdots a_{|u|-1}$ is finite, we have $u[i, j] = a_i \cdots a_j$.

(2) Otherwise, $u[i, j] = \varepsilon$.

Languages and classes. A *language* over an alphabet A is a subset of A^*. Similarly, an *ω-language* is a subset of A^ω. In the paper, we investigate classes of languages and classes of ω-languages. A *class of languages* C is a map $A \mapsto C(A)$ associating a set $C(A)$ of languages over A to each alphabet A. Similarly, a *class of ω-languages* is a map $A \mapsto C(A)$ which associates a set $C(A)$ of ω-languages over A to each alphabet A.

REMARK 2.1. *For the sake of simplifying the presentation, it is usual to abuse notation by making the alphabet implicit: when A is clear from the context, one simply writes $L \in C$ for $L \in C(A)$. Note however that we shall often manipulate distinct alphabets simultaneously.*

In the paper, we work with regular languages. The regular languages are those that can be equivalently defined by *nondeterministic finite automata* (NFA), *finite monoids* or *monadic second-order logic* (MSO) interpreted on words. Similarly, regular ω-languages are those that can be equivalently defined by *nondeterministic Büchi automata* (NBA), *finite ω-semigroups* or MSO interpreted on ω-words. In the paper we work with the algebraic definition of regular languages and ω-languages in terms of monoids and ω-semigroups. We recall these notions in Sections 4 and 7 respectively.

2.2 Closure properties

In the paper, we only consider classes satisfying robust closure properties that we present now. We define them for classes of languages (the corresponding definitions for ω-languages are analogous).

Boolean operations. We only consider lattices. A *lattice of languages* is a class of languages C such that for any alphabet A, the two following properties are satisfied:

- *Closure under union.* For any $L_1, L_2 \in C(A)$, we have $L_1 \cup L_2 \in C(A)$. Moreover, $C(A)$ contains the empty union: $\emptyset \in C(A)$.
- *Closure under intersection.* For any $L_1, L_2 \in C(A)$, we have $L_1 \cap L_2 \in C(A)$. Moreover, $C(A)$ contains the empty intersection: $A^* \in C(A)$.

A *Boolean algebra of languages* is a lattice closed under complement: for any alphabet A, if $L \in C(A)$ then $A^* \setminus L \in C(A)$.

REMARK 2.2. *Note that since ω-languages are subsets of A^ω, the empty intersection and complement are interpreted over A^ω for classes of ω-languages. For example, the empty intersection is A^ω, for any alphabet A.*

Quotient. We shall also consider closure under *right quotient* (we do not need left quotient). Consider an alphabet A. Given $L \subseteq A^*$ and any $u \in A^*$, we define the right quotient $Lu^{-1} \subseteq A^*$ of L by u as the language,

$$Lu^{-1} \stackrel{\text{def}}{=} \{w \in A^* \mid wu \in L\}.$$

We say that a class of languages C is *closed under right quotient* when for any alphabet, any $L \in C(A)$ and any $u \in A^*$, $u^{-1}L \in C(A)$. We shall not consider closure under quotient for classes of ω-languages.

Inverse image. Finally, we also consider closure under inverse image. For the definition, we need to introduce monoid morphisms. A *semigroup* is a set S equipped with an associative multiplication, written $s \cdot t$ or st. A *monoid* is a semigroup M having a neutral element 1_M, i.e., such that $s \cdot 1_M = 1_M \cdot s = s$ for all $s \in M$. Moreover, a monoid morphism is a mapping $\alpha : M \to N$ from a monoid to another, which respects the algebraic structure: for all $s, s' \in M$, we have $\alpha(s \cdot s') = \alpha(s) \cdot \alpha(s')$ and $\alpha(1_M) = 1_N$. Observe that for any alphabet A, the sets A^+ and A^* are respectively a semigroup and a monoid when equipped with concatenation (the neutral element of A^* is ε). Therefore, given any two alphabets A, B, we may define morphisms $\alpha : A^* \to B^*$.

Given a class of languages C, we say that C is *closed under inverse image* when for any two alphabets A, B, any morphism $\alpha : A^* \to B^*$ and any language $L \in C(B)$, we have $\alpha^{-1}(L) \in C(A)$. We shall also consider a weaker variant of closure under inverse image: *alphabetic* inverse image. We say that a morphism $\alpha : A^* \to B^*$ is *alphabetic* when $\alpha(a) \in B$ for any letter $a \in A$ (the image of a letter is a letter). A class of languages C is *closed under alphabetic inverse image* when for any two alphabets A, B, any *alphabetic* morphism $\alpha : A^* \to B^*$ and any language $L \in C(B)$, we have $\alpha^{-1}(L) \in C(A)$.

We finish by lifting the definition of inverse image to classes of ω-languages. Observe we may lift any morphism $\alpha : A^* \to B^*$ as a map $\alpha : A^\infty \to B^\infty$. Indeed, if $w \in A^*$, then $\alpha(w)$ is already defined and if $w = a_0 a_1 a_2 \cdots \in A^\omega$, then we may define,

$$\alpha(w) = \alpha(a_0)\alpha(a_1)\alpha(a_2)\cdots \in B^\infty.$$

REMARK 2.3. *Note that when $w \in A^\omega$, $\alpha(w)$ may belong to either B^ω or B^*. This depends on whether there are infinitely many indices $n \in \mathbb{N}$ such that $\alpha(a_n) \neq \varepsilon$. On the other hand, given an ω-language $L \subseteq B^\omega$, its inverse image $\alpha^{-1}(L)$ is necessarily an ω-language as well, i.e., a subset of A^ω.*

Given a class of ω-languages C, we say that C is closed under inverse image when for any two alphabets A, B, any map $\alpha : A^\infty \to B^\infty$ generated by a morphism and any ω-language $L \in C(B)$, we have $\alpha^{-1}(L) \in C(A)$.

2.3 Decision problems

We turn to the two decision problems that we shall consider: separation and covering. Both of them are parametrized by an arbitrary class of languages or ω-languages C and their purpose is to serve as mathematical tools for analyzing C. We only present the definition for classes of languages (adapting it to ω-languages is immediate).

Separation. Given three languages K, L_1, L_2, we say that K *separates* L_1 from L_2 if $L_1 \subseteq K$ and $K \cap L_2 = \emptyset$. Furthermore, if C is some class of languages and L_1, L_2 are two languages, we say that L_1 is C-*separable* from L_2 when there exists $K \in C$ that separates L_1 from L_2.

REMARK 2.4. *Observe that when* C *is closed under complement,* L_1 *is* C-*separable from* L_2 *if and only if* L_2 *is* C-*separable from* L_1. *However, this is not true for classes that are not closed under complement.*

Given a class of languages C, we may now define the C-separation problem as follows:

INPUT: Two regular languages L_1 and L_2.
OUTPUT: Is L_1 C-separable from L_2?

When investigating separation for a particular class C, one usually considers two complementary objectives: finding an algorithm that decides it and finding a generic for constructing a separator in C when there exists one.

REMARK 2.5. *Separation generalizes another well-known decision problem:* membership. *Given a class* C, *this problem asks whether an input regular language* L *belongs to* C. *This is equivalent to asking whether it is* C-*separable from its complement (which is also regular). Indeed, in that case, there is only one candidate for being a separator:* L *itself. In other words,* C-*membership reduces to* C-*separation.*

Covering. We now present the covering problem which, we originally introduced in [32, 34] as a natural generalization of separation.

REMARK 2.6. *One of the primary motivations for introducing covering is that even if one is only interested in separation, considering covering is required for many classes.*

Given a language L, a *cover of* L is a **finite** set of languages $\mathbf{K}$ such that $L \subseteq \bigcup_{K \in \mathbf{K}} K$. Moreover, given a class C, a C-*cover of* L is a cover $\mathbf{K}$ of L such that all $K \in \mathbf{K}$ belong to C. Additionally, given a finite multiset[1] of languages $\mathbf{L}$, we say that a finite set of languages $\mathbf{K}$ is *separating* for $\mathbf{L}$ if for any $K \in \mathbf{K}$, there exists $L \in \mathbf{L}$ such that $K \cap L = \emptyset$ (*i.e.*, no element of $\mathbf{K}$ intersects all languages in $\mathbf{L}$).

Consider a class C. Given a language L_1 and a finite multiset of languages $\mathbf{L}_2$, we say that the pair $(L_1, \mathbf{L}_2)$ is C-*coverable* when there exists a C-cover of L_1 which is separating for $\mathbf{L}_2$. The C-*covering problem* is as follows:

INPUT: A regular language L_1 and a finite multiset of regular languages $\mathbf{L}_2$.
OUTPUT: Is $(L_1, \mathbf{L}_2)$ is C-coverable?

As for separation, one has usually two goals when investigating C-covering: getting an algorithm that decides it and finding a generic method for building separating C-covers when they exist. We complete this definition by explaining why covering generalizes separation: the latter is special case of the former when the multiset $\mathbf{L}_2$ is a singleton (provided that the class C is a lattice). We state this in the following fact whose proof is easy and given in [34].

FACT 2.7. *Let* C *be a lattice and* L_1, L_2 *two languages. Then* L_1 *is* C-*separable from* L_2, *if and only if* $(L_1, \{L_2\})$ *is* C-*coverable.*

---

[1]We speak of multiset here for the sake of allowing several copies of the same language in $\mathbf{L}$. This is natural. Indeed, $\mathbf{L}$ is an input of our problem: what we have in hand is a set of *recognizers* for the languages in $\mathbf{L}$, and distinct recognizers may well define the same language.

2.4 Suffix languages

We finish this preliminary section by presenting a specific class of languages: the *suffix languages* (SU). While simple, SU will be crucial in the paper: we use it in a generic construction which builds new classes on top of already existing ones.

We first define SU and then present a classification of the languages it contains. For any alphabet A, SU(A) consists of all finite Boolean combinations of languages of the form A^*w for some $w \in A^*$. It is immediate by definition that SU is a Boolean algebra. We state this in the following proposition.

PROPOSITION 2.8. SU *is a Boolean algebra.*

REMARK 2.9. *While we shall not need this property,* SU *is also closed under quotient. On the other hand, it is not closed under inverse image.*

We also consider the following classification of the languages in SU (we call it a *stratification* of SU). For any $k \in \mathbb{N}$, we define a *finite* class SU_k (*i.e.*, $\mathrm{SU}_k(A)$ is a finite set for any A). Given an alphabet A, $\mathrm{SU}_k(A)$ consists of all finite Boolean combinations of languages having the form A^*w for some $w \in A^*$ such that $|w| \le k$. One may verify that all strata SU_k are finite Boolean algebras. Moreover, for any alphabet A, we have:

$$\mathrm{SU}_k(A) \subseteq \mathrm{SU}_{k+1}(A) \text{ for any } k \in \mathbb{N} \qquad \text{and} \qquad \bigcup_{k \in \mathbb{N}} \mathrm{SU}_k(A) = \mathrm{SU}(A).$$

Our motivation for introducing this stratification of SU is the canonical equivalence that one may associate to each stratum SU_k. Consider an alphabet A. For any natural number $k \in \mathbb{N}$ and any two words $w, w' \in A^*$, we write $w \sim_k w'$ if and only if the following condition holds:

$$\text{For any language } L \in \mathrm{SU}_k(A), \quad w \in L \Leftrightarrow w' \in L.$$

By definition and since SU_k is finite, $\sim_k$ is an equivalence relation of finite index. Finally, since all strata SU_k are Boolean algebras, the following lemma is immediate.

LEMMA 2.10. *Let $k \in \mathbb{N}$. Then for any alphabet A, the languages in $\mathrm{SU}_k(A)$ are exactly the unions of $\sim_k$-classes.*

Finally, we shall need the following result which follows Lemma 2.10 and gives an alternate definition of SU, which is sometimes simpler to manipulate. One may verify that for any $k \in \mathbb{N}$, the equivalence classes of $\sim_k$ are all languages of the form A^*w for $|w| = k$ or $\{w\}$ for some $|w| \le k - 1$. Thus, we have the following lemma.

LEMMA 2.11. *Let A be an alphabet, let $k \in \mathbb{N}$ and let L be a language over A. Then, $L \in \mathrm{SU}_k(A)$ if and only if L is a union of languages having one the two following forms:*

(1) A^*w *for some $w \in A^*$ such that $|w| = k$.*
(2) $\{w\}$ *for some $w \in A^*$ such that $|w| \le k - 1$.*

3 FRAGMENTS OF FIRST-ORDER LOGIC

In this section, we briefly recall the definition of first-order logic over words and ω-words. Moreover, we introduce the various fragments that we intend to investigate.

3.1 First-order logic and fragments

We first briefly define first-order logic. Consider an alphabet A. Recall that we view a word $w \in A^*$ as a linearly ordered set of labeled positions $\{0, \ldots, |w| - 1\}$. In first-order logic (FO(<)), one can quantify over these positions and use the following predicates:

- *The label predicates*: for each $a \in A$, a unary predicate "$a(x)$" selects all positions labeled with an a.
- *Linear order*: a binary predicate "$x < y$" interpreted as the (strict) linear order over the positions.

Each first-order sentence defines the language of all words satisfying it. For example, the sentence "$\exists x \exists y \, (x < y \land a(x) \land b(y))$" defines the language $A^*aA^*bA^*$. We shall freely use the name "FO($<$)" to denote both first-order logic and the class of languages that may be defined by a first-order sentence.

Moreover, FO($<$) also defines a class of ω-languages. Recall that the set of positions in an ω-word is simply $\mathbb{N}$. Thus, we may interpret FO($<$) sentences on ω-words. For example, the sentence "$\exists x \exists y \, (x < y \land a(x) \land b(y))$" also defines the ω-language $A^*aA^*bA^\omega$. Therefore, FO($<$) defines two classes: a class of languages and a class of ω-languages. We speak of FO($<$) over words and FO($<$) over ω-words. We shall adopt a similar terminology for all fragments that we consider.

First-order logic itself is well-understood. The solution to the membership problem for FO($<$) over words is due to Schützenberger [38], McNaughton and Papert [18]. It is considered as a seminal result for this research field. It was later lifted to ω-words by Perrin [19]. Separation and covering were considered much later. Both problems were solved for words and ω-words by the authors [30, 33]. However, the focus of our investigation in the present paper is not first-order logic itself. Instead, we are interested in specific fragments of first-order logic that we define now.

Two-variable first-order logic. This fragment is denoted by FO^2($<$). It restricts FO($<$) sentences to those containing at most **two** distinct variables. Note however that these two variables may be reused. For example, the sentence

$$\exists x \exists y \, x < y \land a(x) \land b(y) \land (\exists x \, y < x \land c(x))$$

is an FO^2($<$) sentence defining the language $A^*aA^*bA^*cA^*$ and the ω-language $A^*aA^*bA^*cA^\omega$. It is folklore and simple to verify that over words, FO^2($<$) is a Boolean algebra closed under quotient and inverse image. Similarly, over ω-words, FO^2($<$) is a Boolean algebra closed under inverse image.

Quantifier alternation hierarchy. We shall also consider fragments within the quantifier alternation of first-order logic. It is natural to classify first-order sentences by counting the number of alternations between $\exists$ and $\forall$ quantifiers in their prenex normal form. More precisely, given a natural number $n \geq 1$, an FO($<$) sentence is $\Sigma_n(<)$ (resp. $\Pi_n(<)$) when its prenex normal form has $(n-1)$ quantifier alternations (that is, n blocks of quantifiers) and starts with an $\exists$ (resp. a $\forall$) quantifier. For example, a sentence whose prenex normal form is

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

is $\Sigma_3(<)$. Observe that the sets of $\Sigma_n(<)$ and $\Pi_n(<)$ sentences are not closed under negation: negating a $\Sigma_n(<)$ sentence yields a $\Pi_n(<)$ sentence and vice versa. Thus, one also considers $\mathcal{B}\Sigma_n(<)$ sentences: Boolean combinations of $\Sigma_n(<)$ sentences. This yields hierarchies of classes of languages and ω-languages, and both are strict [3]. The hierarchy for words is depicted in Figure 1. Colors depict the status of each fragment: green ($\Sigma_1(<)$, $\Sigma_2(<)$, $\Sigma_3(<)$, $\mathcal{B}\Sigma_1(<)$, $\mathcal{B}\Sigma_2(<)$) means that covering is decidable (hence also separation and membership); blue ($\Pi_1(<)$, $\Pi_2(<)$, $\Pi_3(<)$) means that separation is decidable, while the status for covering is unknown; yellow ($\Sigma_4(<)$, $\Pi_4(<)$) means that membership is decidable for words and unknown for ω-words[2], while the status for separation and covering is unknown; finally, red means that even the status for membership is unknown (which is also the case for $\mathcal{B}\Sigma_4(<)$ and all fragments above).

---

[2] Actually, for ω-words, very few results on separation and covering have been published, see Section 7.
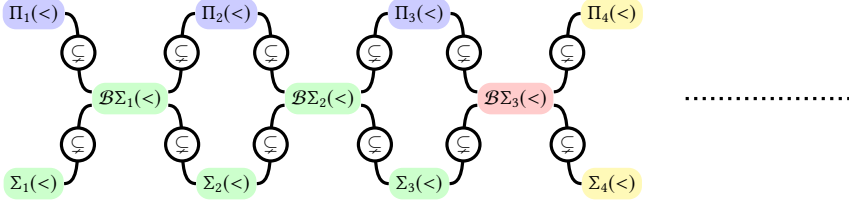
Fig. 1. Quantifier alternation hierarchy within FO(<).

It is folklore and simple to verify that over words, all levels $\Sigma_n(<)$ (resp. $\mathcal{B}\Sigma_n(<)$) are lattices (resp. Boolean algebras) closed under quotient and inverse image. Similarly, over ω-words, all levels $\Sigma_n(<)$ (resp. $\mathcal{B}\Sigma_n(<)$) are lattices (resp. Boolean algebras) closed under inverse image. We shall come back to the quantifier alternation hierarchy of FO(<) in Section 6.

3.2 Enriched signatures

Observe that one may define a seemingly stronger variant FO(<, +1) of FO(<) by enriching its signature with the following natural predicates:

- A binary predicate "$x + 1 = y$" interpreted as the successor relation between positions.
- A unary predicate "$min(x)$" which selects the leftmost position.
- A unary predicate "$max(x)$" which selects the rightmost position (in a finite word).
- A constant "ε" which holds for the empty word.

REMARK 3.1. *Naturally, "max" and "ε" are only useful when interpreting first-order sentences over finite words: ω-words cannot have a rightmost position nor be empty.*

Note however that FO(<, +1) is only stronger in the *syntactic* sense: it is known and simple to verify that FO(<) and FO(<, +1) have the same expressive power, *i.e.*, the corresponding classes of languages (resp. of ω-languages) are the same. In other words, the four above predicates min, max, +1 and ε may be defined from the linear order. For example $x + 1 = y$ is defined by the formula $x < y \land \neg(\exists z \; x < z \land z < y)$.

Nonetheless, this remark is crucial for the paper: we are not interested in first-order logic itself but in its *fragments*. It turns out that for FO^2(<) and the levels $\Sigma_n(<)$, $\mathcal{B}\Sigma_n(<)$ in the quantifier alternation hierarchy, adding the above predicates to the signature yields strictly more expressive logics. Therefore, for each fragment, we are able to define **two** natural "variants": a weak one (whose signature consists of the linear order < and of the letter predicates $a()$) and a strong one (obtained by enriching the signature of the weak one with the above predicates). We write FO^2(<, +1) for the strong variant of FO^2(<) and $\Sigma_n(<, +1)$, $\mathcal{B}\Sigma_n(<, +1)$ for the strong variants of $\Sigma_n(<)$, $\mathcal{B}\Sigma_n(<)$.

Example 3.2. It is not possible to express successor in the two-variables restriction of first-order logic (FO^2(<)). Intuitively, this is because it requires quantifying over a third variable (on the other hand, it is simple to express "min", "max" and"ε").

REMARK 3.3. *While the predicates "min", "max", "ε" are not explicitly mentioned in our notation, they are allowed in all strong variants. We omit them in the notation since they can be defined from "<" for all fragments except for $\Sigma_1(<), \Pi_1(<)$ and $\mathcal{B}\Sigma_1(<)$.*

Our objective in the paper is to investigate the covering and separation problems associated to strong variants. Our main contribution is a *generic reduction technique*. It is designed to exploit the

intuitive relationship between weak and strong variants: we reduce covering for the latter to the same problem for the former. We are then able to obtain covering and separation algorithms for several strong variants as corollaries of already existing results for the corresponding weak variants. In fact, we have two similar reduction theorems: one for finite words (presented in Section 4) and one for ω-words (presented in Section 7).

Let us sketch our approach using the case of finite words. The reduction technique is based on two ingredients. The first one is a generic operation that can be applied to a class of languages: SU-*enrichment* (here, SU denotes the class of suffix languages defined in Section 2). Given a class of languages C, SU-enrichment constructs a larger class denoted $C \circ \mathrm{SU}$. The key idea is that SU-enrichment captures the intuitive relationship between weak and strong variants as a formal and generic connection between the associated classes: if C is the class corresponding to a weak variant, then $C \circ \mathrm{SU}$ corresponds to the strong variant.

REMARK 3.4. *The connection between weak and strong variants is not a new result, even formalized as a generic operation defined on classes of languages. For example, this was observed by Straubing [42] for the quantifier alternation hierarchy. However, an important point is that in the literature, these results are usually formulated using algebraic terminology. In contrast, in this paper, we work directly at the level of language classes. We shall come back to this point in the next section.*

The second ingredient is formulated as a generic reduction theorem. Given a lattice of languages C closed under right quotient and inverse image, it states that $(C \circ \mathrm{SU})$-covering reduces to C-covering. By combining the two ingredients, we get a reduction that is generic to all logical fragments outlined above.

4 REDUCING STRONG TO WEAK VARIANTS

In this section, we present our generic reduction for languages of finite words. First, we define an operation which combines two class of languages C and $\mathcal{D}$ into a larger one $C \circ \mathcal{D}$: the "$\mathcal{D}$-*enrichment of C*". This operation is crucial: as we shall prove later, when $\mathcal{D} = \mathrm{SU}$, it captures the intuitive connection existing between weak and strong logical fragments.

REMARK 4.1. *Enrichment is the language theoretic counterpart of an algebraic operation defined between varieties of semigroups: the* wreath product *(see [42] for words and [7] for ω-words). In fact, we use the same notation: "$\circ$". We have two motivations for using a pure language theoretic point of view here:*

(1) *This is much simpler: we avoid a lot of algebraic machinery.*

(2) *Manipulating this definition in proofs is more natural since we are only dealing with one object:* classes of languages. *On the other hand, using the algebraic definition requires handling varieties of semigroups and classes of languages simultaneously.*

Our approach makes it necessary to prove the connection with logic, i.e., that SU-enrichment captures the link between weak and strong fragments.

In this section, we first define enrichment and then present our reduction theorem: given any class of languages C (satisfying appropriate properties), $(C \circ \mathrm{SU})$-covering reduces to C-covering. The remainder of the section will be devoted to proving this result.

4.1 The enrichment of a class of languages

We are now ready to define enrichment. We start with a preliminary notion.

P-taggings. Let A be an alphabet and let $\mathbf{P}$ be a finite partition of A^*. Observe that $\mathbf{P} \times A$ is also a finite set. We use it as an extended alphabet and define a canonical map $\tau_{\mathbf{P}} : A^* \to (\mathbf{P} \times A)^*$. Given an arbitrary word $u \in A^*$, we denote by $[u]_{\mathbf{P}}$ the unique language in the partition $\mathbf{P}$ that contains u.

Let $w \in A^*$ be a word. Consider the decomposition of w as a concatenation of letters: $w = a_1 \cdots a_n \in A^*$. We let $\tau_{\mathbf{P}}(w)$ be the word $\tau_{\mathbf{P}}(w) = b_1 \cdots b_n \in (\mathbf{P} \times A)^*$ where,

$$b_1 = ([\varepsilon]_{\mathbf{P}}, a_1) \quad \text{and} \quad b_i = ([a_1 \cdots a_{i-1}]_{\mathbf{P}}, a_i) \quad \text{for } 2 \le i \le n.$$

Note that when w is empty, then $\tau_{\mathbf{P}}(\varepsilon) = \varepsilon$. Given any $w \in A^*$, we call $\tau_{\mathbf{P}}(w)$ the $\mathbf{P}$-*tagging* of w. Observe that the $\mathbf{P}$-tagging of w is simply a relabeling: each position i in w is given a new label encoding its original label in A and the unique language in $\mathbf{P}$ containing the prefix $w[1, i-1]$.

Example 4.2. Let $A = \{a, b\}$ and consider the languages $P_\varepsilon = \{\varepsilon\}$, $P_a = A^*a$ and $P_b = A^*b$. Clearly, $\mathbf{P} = \{P_\varepsilon, P_a, P_b\}$ is a partition of A^*. Let $w = babba \in A^*$, the $\mathbf{P}$-tagging of w is $\tau_{\mathbf{P}}(w) = (P_\varepsilon, b)(P_b, a)(P_a, b)(P_b, b)(P_b, a)$.

REMARK 4.3. *The map $w \mapsto \tau_{\mathbf{P}}(w)$ is not a morphism. Moreover, it is not surjective in general. Indeed, there are usually compatibility constraints between consecutive positions in $\tau_{\mathbf{P}}(w)$, as can be observed in Example 4.2 (on the other hand, $\tau_{\mathbf{P}}$ is injective).*

While $\tau_{\mathbf{P}}$ is not a morphism, it will often be convenient to decompose the image $\tau_{\mathbf{P}}(w)$ of some word $w \in A^*$. For this, we shall use a second map $\delta_{\mathbf{P}} : A^* \times A^* \to A^*$. Let $u, w \in A^*$ and consider the decomposition of w as a concatenation of letters: $w = a_1 \cdots a_n \in A^*$. We let $\delta_{\mathbf{P}}(u, w) = b_1 \cdots b_n \in (\mathbf{P} \times A)^*$ where,

$$b_1 = ([u]_{\mathbf{P}}, a_1) \quad \text{and} \quad b_i = ([ua_1 \cdots a_{i-1}]_{\mathbf{P}}, a_i) \quad \text{for } 2 \le i \le n.$$

The following lemma may be verified from the definitions of $\tau_{\mathbf{P}}$ and $\delta_{\mathbf{P}}$.

LEMMA 4.4. *Let A be an alphabet, $\mathbf{P}$ a finite partition of A^*. Then given any $u \in A^*$ and any $w \in A^*$, we have $\tau_{\mathbf{P}}(uw) = \tau_{\mathbf{P}}(u) \cdot \delta_{\mathbf{P}}(u, w)$.*

Definition of enrichment. Consider two classes of languages C and $\mathcal{D}$. We define a new class of languages $C \circ \mathcal{D}$ called the $\mathcal{D}$-*enrichment of C*. Given an alphabet A, a $\mathcal{D}$-*partition of A^** is a *finite* partition of A^* into languages of $\mathcal{D}$.

We define $C \circ \mathcal{D}$ as the following class of languages. Let A be some alphabet. Given any language $L \subseteq A^*$, we have $L \in (C \circ \mathcal{D})(A)$ if and only if there exists a $\mathcal{D}$-partition $\mathbf{P}$ of A^* and languages $L_P \in C(\mathbf{P} \times A)$ for all $P \in \mathbf{P}$ such that,

$$L = \bigcup_{P \in \mathbf{P}} \left(P \cap \tau_{\mathbf{P}}^{-1}(L_P) \right).$$

REMARK 4.5. *Strictly speaking, this definition makes sense for any two classes of languages C and $\mathcal{D}$. On the other hand, one needs a few hypotheses for it to be robust. Specifically, it is natural to require C to be closed under alphabetic inverse image: since we use C for distinct alphabets, it makes sense to have a property connecting what C is for these distinct alphabets. Similarly, requiring $\mathcal{D}$ to be a Boolean algebra is natural since we use $\mathcal{D}$-partitions of A^*.*

It turns out that when C is a lattice of languages closed under alphabetic inverse image and $\mathcal{D}$ is a Boolean algebra of languages, $C \circ \mathcal{D}$ is a lattice of languages containing both C and $\mathcal{D}$ (since we never use this property, its proof is left to the reader).

Example 4.6. Consider the class AT of alphabet testable languages: for any alphabet A, AT(A) contains all Boolean combinations of languages A^*aA^* for $a \in A$. We describe a language in

AT ∘ SU. Let $A = \{a, b\}$. We claim that A^*abA^*a belongs to $(\text{AT} \circ \text{SU})(A)$. Consider the SU-partition of Example 4.2: $\mathbf{P} = \{P_\varepsilon, P_a, P_b\}$ where $P_\varepsilon = \{\varepsilon\}$, $P_a = A^*a$ and $P_b = A^*b$. Clearly, $L = (\mathbf{P} \times A)^* \cdot (P_a, b) \cdot (\mathbf{P} \times A)^*$ belongs to $\text{AT}(\mathbf{P} \times A)$. Moreover, one may verify that:

$$A^*abA^*a = (P_\varepsilon \cap \tau_\mathbf{P}^{-1}(\emptyset)) \cup (P_a \cap \tau_\mathbf{P}^{-1}(L)) \cup (P_b \cap \tau_\mathbf{P}^{-1}(\emptyset)) \in \text{AT} \circ \text{SU}.$$

In fact, AT ∘ SU is exactly the class of locally testable languages, which is well-known in the literature [4, 17, 47].

REMARK 4.7. *If the class $\mathcal{D}$ can be written $\mathcal{D} = \bigcup_{k \in \mathbb{N}} \mathcal{D}_k$, then we obtain from the definition that $C \circ \mathcal{D} = \bigcup_{k \in \mathbb{N}} C \circ \mathcal{D}_k$. In the case of SU-enrichment, we will use this remark with the natural stratification $(\text{SU}_k)_{k \in \mathbb{N}}$ of SU defined in Section 2.4.*

As we explained, we are mainly interested in the special case of SU-enrichment as it captures the intuitive connection between strong and weak logical fragments. Specifically, we prove in Section 5 that $\text{FO}^2(<, +1)$ is the SU-enrichment of $\text{FO}^2(<)$. Moreover, we show in Section 6 that for any $n \geq 1$ $\Sigma_n(<, +1)$ and $\mathcal{B}\Sigma_n(<, +1)$ are respectively the SU-enrichments of $\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<)$.

We now turn to the main theorem of the paper (more precisely, to its variant for finite words): given any lattice of languages C closed under *right quotient* and *inverse image*, $(C \circ \text{SU})$-covering reduces to C-covering. The remainder of the section is devoted to presenting this reduction. Let us start with an outline of the different steps it involves.

The reduction works with the algebraic definition of regular languages, which we recall. As we explained, given any alphabet A, the universal language A^* is a monoid. Given an arbitrary monoid M and a language $L \subseteq A^*$, we say that L is *recognized by M* if there exists a monoid morphism $\alpha : A^* \to M$ and a set $F \subseteq M$ such that $L = \alpha^{-1}(F)$. It is well-know that a language is regular if and only if it can be recognized by a *finite* monoid. Moreover, if L is regular, one may compute such a morphism recognizing L from any representation of L (such as an NFA or an MSO sentence).

Consider an input pair $(L, \mathbf{L})$ for the covering problem: L is a regular language and $\mathbf{L}$ is a finite multiset of regular languages. Our reduction requires starting from a single monoid morphism $\alpha : A^* \to M$ recognizing all languages in $\{L\} \cup \mathbf{L}$. This is mandatory, as the reduction is parametrized by α.

REMARK 4.8. *This requirement is not restrictive: it is simple to build such a morphism. Assume that $\{L\} \cup \mathbf{L} = \{L_1, \ldots, L_n\}$. For each $i \leq n$, we may build a morphism $\alpha_i : A^* \to M_i$ recognizing L_i. It then suffices to use the morphism $\alpha : A^* \to M_1 \times \cdots \times M_n$ defined by $\alpha(w) = (\alpha_1(w), \ldots, \alpha_n(w))$, which recognizes all languages L_i.*

Once we have the morphism $\alpha : A^* \to M$ in hand, we use it in a generic construction which builds two objects. The first one is a new alphabet $\mathbb{A}_\alpha$: the *alphabet of well-formed words*. The second one is a map $L \mapsto \text{wf}_\alpha(L)$ which associates a new regular language over $\mathbb{A}_\alpha$ to any language $L \subseteq A^*$ recognized by α. We also extend this map to multisets.

Assume that C is a lattice of languages closed under right quotient and inverse image. The reduction states that for any pair $(L, \mathbf{L})$ such that all languages in $\{L\} \cup \mathbf{L}$ are recognized by α, the two following properties are equivalent:

(1) $(L, \mathbf{L})$ is $(C \circ \text{SU})$-coverable.
(2) $(\text{wf}_\alpha(L), \text{wf}_\alpha(\mathbf{L}))$ is C-coverable.

This concludes our outline. The remainder of this section is organized as follows. We first present the construction which builds an alphabet of well-formed words from an arbitrary morphism. Then, we state the reduction theorem and prove it.

4.2   Languages of well-formed words

We describe a generic construction which takes as input a morphism $\alpha : A^* \to M$ into a finite monoid M. It builds the following objects:

(1) An alphabet $\mathbb{A}_\alpha$, called the *alphabet of well-formed words associated to α*.
(2) A map associating to any language L over A^* *recognized by α* a regular language $\text{WF}_\alpha(L)$ over $\mathbb{A}_\alpha$. We call $\text{WF}_\alpha(L)$ the *language of well-formed words associated to L*.

We denote by S the semigroup $S = \alpha(A^+)$, that is, the image in M of all **nonempty** words. Moreover, we write $E(S)$ for the set of idempotent elements in S (*i.e.*, $E(S)$ consists of all $e \in S$ such that $ee = e$). We also write S^1 for $S \cup \{1_M\}$ (notice that it may happen that $S = S^1$). Finally, we let "$\square$" be some symbol that does not belong to M. The *alphabet of well-formed words associated to α*, denoted by $\mathbb{A}_\alpha$, is defined as follows:

$$\mathbb{A}_\alpha = (E(S) \cup \{\square\}) \times S^1 \times (E(S) \cup \{\square\}).$$

Note that since S depends on α, so does $\mathbb{A}_\alpha$. We are not interested in all words of $\mathbb{A}_\alpha^*$, but only in those that are "well-formed". Given a word $w \in \mathbb{A}_\alpha^*$, we say that w is *well-formed* if it is **nonempty** (*i.e.*, $w \in \mathbb{A}_\alpha^+$) and has the following form:

$$w = (\square, s_1, f_1) \cdot (e_2, s_2, f_2) \cdots (e_{n-1}, s_{n-1}, f_{n-1}) \cdot (e_n, s_n, \square)$$

with $f_i = e_{i+1} \in E(S)$ for all $1 \le i \le n-1$. In other words,

$$w = (\square, s_1, e_2) \cdot (e_2, s_2, e_3) \cdots (e_{n-1}, s_{n-1}, e_n) \cdot (e_n, s_n, \square).$$

In particular, well-formed words of length 1 are of the form $(\square, s, \square)$ with $s \in S^1$.

REMARK 4.9. *The definition requires that $f_i = e_{i+1} \in E(S)$ for all $1 \le i \le n-1$. This means that the idempotents $e_2, \ldots, e_n$ must be the image under α of a **nonempty** word. This is easy to miss. However, this property is crucial for proving the main theorem.*

It is straightforward to build an automaton recognizing the language of well-formed words over $\mathbb{A}_\alpha$. Thus, the following simple fact is immediate.

FACT 4.10. *The language of all well-formed words over $\mathbb{A}_\alpha$ is regular.*

We now associate a new language over $\mathbb{A}_\alpha$ to each language L recognized by α: we call it the *language of well-formed words associated to L*. As the name suggests, it is made exclusively of well-formed words.

One defines a canonical morphism $\text{EVAL} : \mathbb{A}_\alpha^* \to M$ by defining the image of each letter from $\mathbb{A}_\alpha$ (there are four kinds of such letters). For $s \in S^1$ and $e, f \in E(S)$, we let:

$$\begin{cases} \text{EVAL}((e, s, f)) &=& esf, \qquad \text{EVAL}((\square, s, f)) &=& sf, \\ \text{EVAL}((e, s, \square)) &=& es, \qquad \text{EVAL}((\square, s, \square)) &=& s. \end{cases}$$

Consider a language L recognized by α. The *language of well-formed words associated to L*, denoted by $\text{WF}_\alpha(L)$, is defined as follows:

$$\text{WF}_\alpha(L) = \left\{ w \in \mathbb{A}_\alpha^* \mid w \text{ is well-formed and } \text{EVAL}(w) \in \alpha(L) \right\} \subseteq \mathbb{A}_\alpha^+.$$

Observe that by definition, $\text{WF}_\alpha(L)$ is the intersection of some language recognized by EVAL with the language of well-formed words. Hence, we have the following fact.

FACT 4.11. *For any language $L \subseteq A^*$ recognized by α, $\text{WF}_\alpha(L) \subseteq \mathbb{A}_\alpha^*$ is regular.*

Finally, as explained in the outline, we extend the notation $\text{WF}_\alpha()$ to multisets, by setting $\text{WF}_\alpha(\mathbf{L}) = \{\text{WF}_\alpha(L) \mid L \in \mathbf{L}\}$ for any multiset $\mathbf{L}$ consisting of languages recognized by α.

4.3 Main theorem: reducing $(C \circ \mathrm{SU})$-covering to C-covering

We may now state the theorem that reduces $(C \circ \mathrm{SU})$-covering to C-covering. It is restricted to classes of languages C that are nontrivial: that is, there should exist some alphabet A such that $C(A)$ contains a language L which is neither empty nor universal ($L \neq \emptyset$ and $L \neq A^*$).

THEOREM 4.12 (REDUCTION THEOREM). *Let $\alpha : A^* \to M$ be a morphism and let C be a nontrivial lattice of languages closed under right quotient and inverse image. Moreover, let L be a language and let $\mathbf{L}$ be a multiset of languages, all recognized by α. Then, the following properties are equivalent:*

(1) $(L, \mathbf{L})$ is $(C \circ \mathrm{SU})$-coverable.

(2) $(L, \mathbf{L})$ is $(C \circ \mathrm{SU}_{2|M|})$-coverable.

(3) $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$ is C-coverable.

Before we prove Theorem 4.12, let us discuss its consequences. As announced, the theorem yields a generic reduction from $(C \circ \mathrm{SU})$-covering to C-covering for any nontrivial lattice of languages C closed under right quotient and inverse image.

An important remark is that the proof is constructive. Since we intend to use the theorem as a reduction, this is of particular interest for the direction $(3) \Rightarrow (2)$. It is proved by exhibiting a generic construction: given as input a separating C-cover of $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$, we explain how to build a separating $(C \circ \mathrm{SU}_{2|M|})$-cover of $(L, \mathbf{L})$. Thus, we actually get reductions for the two objectives associated to the covering problem: getting an algorithm that decides it, and finding a generic construction for building separating covers.

Note that one may adapt the statement of Theorem 4.12 to accommodate natural restrictions of covering, such as separation. Recall that this is the special case of inputs $(L, \mathbf{L})$ where $\mathbf{L}$ is a singleton. Thus, we have the following immediate corollary.

COROLLARY 4.13. *Let $\alpha : A^* \to M$ be a morphism and C be a nontrivial lattice of languages closed under right quotient and inverse image. Moreover, let L_1, L_2 be two languages recognized by α. Then, the following properties are equivalent:*

(1) L_1 is $(C \circ \mathrm{SU})$-separable from L_2.

(2) L_1 is $(C \circ \mathrm{SU}_{2|M|})$-separable from L_2.

(3) $\mathrm{WF}_\alpha(L_1)$ is C-separable from $\mathrm{WF}_\alpha(L_2)$.

REMARK 4.14. *In practice, applying Theorem 4.12 to obtain an actual covering or separation algorithm for a given class of languages requires clearing the following preliminary steps:*

(1) Prove that this class is the SU-enrichment $C \circ \mathrm{SU}$ of some lattice of languages C closed under right quotient and inverse image.

(2) Obtain a covering or separation algorithm for C.

The main point here is that it is usually much simpler to achieve these steps than to obtain directly a covering algorithm for the class. For all examples we shall present, we only take care of the first item and obtain the second from previously known results.

We shall apply Theorem 4.12 to obtain covering and separation algorithms for concrete classes of languages in the next two sections.

We devote the rest of this section to proving Theorem 4.12. We keep our notation: $\alpha : A^* \to M$ is a monoid morphism, S is the semigroup $\alpha(A^+)$, and $S^1 = S \cup \{1_M\}$. Recall that the associated alphabet of well-formed words is:

$$\mathbb{A}_\alpha = (E(S) \cup \{\square\}) \times S^1 \times (E(S) \cup \{\square\}).$$

Consider a nontrivial lattice of languages C closed under right quotient and inverse image. Our objective is to show that when L and $\mathbf{L}$ are respectively a language and a multiset of languages recognized by α, the following properties are equivalent:

(1) $(L, \mathbf{L})$ is $(C \circ \mathrm{SU})$-coverable.

(2) $(L, \mathbf{L})$ is $(C \circ \mathrm{SU}_{2|M|})$-coverable.

(3) $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$ is C-coverable.

We prove that $(1) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1)$. Observe that the direction $(2) \Rightarrow (1)$ is trivial since $C \circ \mathrm{SU}_{2|M|} \subseteq C \circ \mathrm{SU}$. Thus, we concentrate on proving that $(1) \Rightarrow (3)$ and $(3) \Rightarrow (2)$.

4.4   From $(C \circ \mathrm{SU})$-covering to C-covering

We start with the implication $(1) \Rightarrow (3)$ in Theorem 4.12. The argument is based on the following proposition.

PROPOSITION 4.15. *For any $k \geq 1$, there exists a map $\gamma : \mathbb{A}_\alpha^* \to A^*$ satisfying the two following properties:*

(1) *For any $L \subseteq A^*$ recognized by α and any well-formed word $w \in \mathbb{A}_\alpha^+$, we have $w \in \mathrm{WF}_\alpha(L)$ if and only if $\gamma(w) \in L$.*

(2) *For any language $K \in (C \circ \mathrm{SU}_k)(A)$, there exists $H_K \in C(\mathbb{A}_\alpha)$ such that for any well-formed word $w \in \mathbb{A}_\alpha^+$, we have $w \in H_K$ if and only if $\gamma(w) \in K$.*

Before proving Proposition 4.15, we use it to show $(1) \Rightarrow (3)$ in Theorem 4.12. Consider a language L and a multiset of languages $\mathbf{L}$, all recognized by α. Assume that $(L, \mathbf{L})$ is $(C \circ \mathrm{SU})$-coverable. We have to prove that $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$ is C-coverable. By hypothesis on $(L, \mathbf{L})$, we have a separating $(C \circ \mathrm{SU})$-cover $\mathbf{K}_A$ for $(L, \mathbf{L})$. We use it together with Proposition 4.15 to construct a separating C-cover $\mathbf{K}_{\mathbb{A}_\alpha}$ for $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$.

Since $\mathbf{K}_A$ contains finitely many languages all belonging to $C \circ \mathrm{SU} = \bigcup_{k \in \mathbb{N}} C \circ \mathrm{SU}_k$ (by Remark 4.7), there is some $k \geq 1$ such that $\mathbf{K}_A \subseteq C \circ \mathrm{SU}_k$. Together with Proposition 4.15, this integer k defines a map $\gamma : \mathbb{A}_\alpha^* \to A^*$. In particular, for any $K \in \mathbf{K}_A$, Item (2) of Proposition 4.15 yields a language $H_K \in C(\mathbb{A}_\alpha)$. We define

$$\mathbf{K}_{\mathbb{A}_\alpha} = \{H_K \mid K \in \mathbf{K}_A\}.$$

To conclude the proof, we show that $\mathbf{K}_{\mathbb{A}_\alpha}$ is a separating C-cover for $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$.

We first prove that $\mathbf{K}_{\mathbb{A}_\alpha}$ is a C-cover of $\mathrm{WF}_\alpha(L)$. Let $w \in \mathrm{WF}_\alpha(L)$, we have to find $H \in \mathbf{K}_{\mathbb{A}_\alpha}$ such that $w \in H$. Since w is well-formed, we know that $\gamma(w) \in L$ by the first item in the proposition. Since $\mathbf{K}_A$ is a cover of L, we can find $K \in \mathbf{K}_A$ such that $\gamma(w) \in K$. It then follows from the second item in the proposition that $w \in H_K$, which belongs to $\mathbf{K}_{\mathbb{A}_\alpha}$ by definition. We conclude that $\mathbf{K}_{\mathbb{A}_\alpha}$ is a cover of $\mathrm{WF}_\alpha(L)$. Moreover, it is a C-cover since all languages in $\mathbf{K}_{\mathbb{A}_\alpha}$ belong to C by Item (2) of Proposition 4.15.

It remains to prove that $\mathbf{K}_{\mathbb{A}_\alpha}$ is separating. Given any $H \in \mathbf{K}_{\mathbb{A}_\alpha}$, we have to find $\mathrm{WF}_\alpha(L') \in \mathrm{WF}_\alpha(\mathbf{L})$ such that $H \cap \mathrm{WF}_\alpha(L') = \emptyset$. By definition, $H = H_K$ for some $K \in \mathbf{K}_A$. Moreover, since $\mathbf{K}_A$ is a separating cover of $(L, \mathbf{L})$, there exists $L' \in \mathbf{L}$ such that $K \cap L' = \emptyset$. This entails that $H_K \cap \mathrm{WF}_\alpha(L') = \emptyset$. Indeed, otherwise, we would have $w \in H_K \cap \mathrm{WF}_\alpha(L')$ which would imply that $\gamma(w) \in K \cap L'$ by the two items of Proposition 4.15, a contradiction. This terminates the proof of $(1) \Rightarrow (3)$ in Theorem 4.12.

It now remains to prove Proposition 4.15.

Proof of Proposition 4.15: definition of γ. Fix a natural number k. We start by defining the map $\gamma : \mathbb{A}_\alpha^* \to A^*$ and we then show that it satisfies the desired properties. It turns out that γ is a morphism. Hence, it suffices describe the image of letters in $\mathbb{A}_\alpha$.

To any element $s \in \alpha(A^*)$, we associate an arbitrarily chosen word $\lceil s \rceil \in A^*$ such that $\alpha(\lceil s \rceil) = s$. When $s \in S = \alpha(A^+)$, we require $\lceil s \rceil$ to be **nonempty** (note that this implies that $\lceil e \rceil \neq \varepsilon$ when $e \in E(S)$, which is crucial in the proof). We are now ready to define our morphism $\gamma : \mathbb{A}_\alpha^* \to A^*$, by defining the image of all four kinds of letters in $\mathbb{A}_\alpha$. Given $s \in S$ and $e, f \in E(S)$, we define,

$$\begin{cases} \gamma((e, s, f)) & = & \lceil e \rceil^k \lceil s \rceil \lceil f \rceil^k, \\ \gamma((\square, s, f)) & = & \lceil s \rceil \lceil f \rceil^k, \\ \gamma((e, s, \square)) & = & \lceil e \rceil^k \lceil s \rceil, \\ \gamma((\square, s, \square)) & = & \lceil s \rceil. \end{cases}$$

Now that we defined the morphism $\gamma : \mathbb{A}_\alpha^* \to A^*$, it remains to prove that it satisfies the two properties of Proposition 4.15. We start with the first one, which is simpler.

Proof of Proposition 4.15: first item. Consider a language $L \subseteq A^*$ recognized by α. We have to show that $w \in \text{WF}_\alpha(L)$ iff $\gamma(w) \in L$, for any well-formed word $w \in \mathbb{A}_\alpha^+$.

Since w is well-formed, $w \in \text{WF}_\alpha(L)$ if and only if $\text{EVAL}(w) \in \alpha(L)$. Moreover, since α recognizes L, $\gamma(w) \in L$ if and only if $\alpha(\gamma(w)) \in \alpha(L)$. Hence, it suffices to show $\text{EVAL}(w) = \alpha(\gamma(w))$. By definition,

$$w = (\square, s_0, e_1) \cdot (e_1, s_1, e_2) \cdots (e_{n-1}, s_{n-1}, e_n) \cdot (e_n, s_n, \square),$$

$$\gamma(w) = \lceil s_0 \rceil \lceil e_1 \rceil^{2k} \lceil s_1 \rceil \lceil e_2 \rceil^{2k} \cdots \lceil e_{n-1} \rceil^{2k} \lceil s_{n-1} \rceil \lceil e_n \rceil^{2k} \lceil s_n \rceil.$$

Hence, we have:

$$\text{EVAL}(w) = s_0 e_1 e_1 s_1 e_2 \cdots e_{n-1} s_{n-1} e_n e_n s_n,$$

$$\alpha(\gamma(w)) = s_0 (e_1)^{2k} s_1 (e_2)^{2k} \cdots (e_{n-1})^{2k} s_{n-1} (e_n)^{2k} s_n.$$

Since the $e_i \in E(S)$ are idempotents, we obtain indeed $\text{EVAL}(w) = \alpha(\gamma(w))$.

Proof of Proposition 4.15: second item. For the proof, we fix some arbitrary language $K \in (C \circ \text{SU}_k)(A)$. We have to build a language $H_K \in C(\mathbb{A}_\alpha)$ such that:

$$\text{For any well-formed word } w \in \mathbb{A}_\alpha^+, \quad w \in H_K \text{ if and only if } \gamma(w) \in K. \tag{1}$$

This is more involved. Recall that by definition of $C \circ \text{SU}_k$, we have an SU_k-partition **P** of A^* and languages $L_P \in C(\mathbf{P} \times A)$ for all $P \in \mathbf{P}$ such that:

$$K = \bigcup_{P \in \mathbf{P}} \left(P \cap \tau_{\mathbf{P}}^{-1}(L_P) \right). \tag{2}$$

Since $C(\mathbb{A}_\alpha)$ is a lattice, it suffices to treat only two particular cases:

- $K \in \mathbf{P}$, and
- $K = \tau_{\mathbf{P}}^{-1}(L_P)$ for some $P \in \mathbf{P}$.

Indeed, if for each $P \in \mathbf{P}$, we are able to exhibit $H_P, H_P' \in C(\mathbb{A}_\alpha)$ such that for any well-formed word $w \in \mathbb{A}_\alpha^+$, we have $w \in H_P$ iff $\gamma(w) \in P$ and $w \in H_P'$ iff $\gamma(w) \in \tau_{\mathbf{P}}^{-1}(L_P)$, then for K given by (2), one can choose $K_H = \bigcup_{P \in \mathbf{P}} (H_P \cap H_P')$. We therefore treat these two cases.

Case 1: $K \in \mathbf{P}$. Therefore, $K \in \text{SU}_k(A)$, since **P** is an SU_k-partition of A^*. In this case, the argument is based on the following fact, which follows from the definition of γ.

FACT 4.16. *Consider two well-formed words $w, w' \in \mathbb{A}_\alpha^+$ with the same rightmost letter. Then, $\gamma(w) \in K$ if and only if $\gamma(w') \in K$.*

PROOF. Since $K \in \text{SU}_k(A)$, it follows from the definition of SU_k that given $u \in A^*$, whether $u \in K$ depends only on the suffixes of length at most k in u. Moreover, by definition of the map γ, given a *well-formed* word $w \in \mathbb{A}_\alpha^+$, the suffixes of length at most k in $\gamma(w)$ depend only on the rightmost letter in w. The fact is then immediate. □

In view of Fact 4.16, there exist a sub-alphabet $\mathbb{B} \subseteq \mathbb{A}_\alpha$ such that for any well-formed word $w \in \mathbb{A}_\alpha^+$, we have $\gamma(w) \in K$ if and only if the rightmost letter in w belongs to $\mathbb{B}$. Thus, it suffices to define a language $H_K \in C(\mathbb{A}_\alpha)$ such that:

$$\text{For any well-formed word } w \in \mathbb{A}_\alpha^+, \quad w \in H_K \text{ if and only if } w \in \mathbb{A}_\alpha^* \cdot \mathbb{B}.$$

It will then be immediate that this language H_K satisfies (1) as desired. It remains to construct $H_K \in C(\mathbb{A}_\alpha)$ satisfying the above property.

For ensuring the condition $H_K \in C(\mathbb{A}_\alpha)$, we use the fact that C is nontrivial. Indeed, this yields an alphabet D such that $C(D)$ contains some language L satisfying $L \neq \emptyset$ and $L \neq D^*$. In particular, we have two words $u, v \in D^*$ such that $u \in L$ and $v \notin L$. Consider the morphism $\eta : \mathbb{A}_\alpha^* \to D^*$ defined as follows. For any letter $b \in \mathbb{A}_\alpha$:

- If b is of the form $(e, s, \square)$ with $s \in S^1$ and $e \in E(S) \cup \{\square\}$ (*i.e.*, b is used as a rightmost letter in some well-formed word), then we define,

$$\eta(b) = \begin{cases} u & \text{if } b \in \mathbb{B}, \\ v & \text{if } b \notin \mathbb{B}. \end{cases}$$

- Otherwise, $\eta(b) = \varepsilon$.

We define $H_K = \eta^{-1}(L)$. Clearly, $H_K \in C(\mathbb{A}_\alpha)$ since C is closed under inverse image. It remains to show that it satisfies the desired property. Let $w \in \mathbb{A}_\alpha^+$ be a well-formed word. By definition of well-formed words, $w = w'b$ where b is the unique letter in w of the form $(e, s, \square)$ with $s \in S^1$ and $e \in E(S) \cup \{\square\}$. Thus, it follows that $\eta(w) = u \in L$ if $b \in \mathbb{B}$ and $\eta(w) = v \notin L$ otherwise. This exactly says that $w \in H_K$ if and only if $w \in \mathbb{A}_\alpha^* \cdot \mathbb{B}$, which concludes the proof of this case.

Case 2: $K = \tau_\mathbf{P}^{-1}(L_P)$ for some $P \in \mathbf{P}$. In this case, the construction of H_K is based on the following lemma.

LEMMA 4.17. *There exists a morphism $\beta : \mathbb{A}_\alpha^* \to (\mathbf{P} \times A)^*$ such that for any well-formed word $w \in \mathbb{A}_\alpha^+$, we have $\tau_\mathbf{P}(\gamma(w)) = \beta(w)$.*

Again, Before proving Lemma 4.17, we use it to construct H_K and finish the proof of Proposition 4.15. We have a language $L_P \in C(\mathbf{P} \times A)$ such that $K = \tau_\mathbf{P}^{-1}(L_P)$. We define

$$H_K = \beta^{-1}(L_P).$$

Since C is closed inverse image, it is immediate from the definition that $H_K \in C(\mathbb{A}_\alpha)$. We now prove that H_K satisfies (1): for any well-formed word $w \in \mathbb{A}_\alpha^+$, $w \in H_K$ if and only if $\gamma(w) \in K$. We use Lemma 4.17. Given a well-formed word $w \in \mathbb{A}_\alpha^+$, we have $w \in H_K$ if and only if $\beta(w) \in L_P$. The lemma then says that this is equivalent to $\tau_\mathbf{P}(\gamma(w)) \in L_P$, *i.e.*, to $\gamma(w) \in K$ by hypothesis on K.

It remains to prove Lemma 4.17. Let us first define the morphism $\beta : \mathbb{A}_\alpha^* \to (\mathbf{P} \times A)^*$. We use the map $\delta_\mathbf{P} : A^* \times A^* \to (\mathbf{P} \times A)^*$ that we defined at the beginning of the section.

Given any letter $(e, s, f) \in \mathbb{A}_\alpha$, we define its image $\beta((e, s, f))$. There are two cases depending on whether $e = \square$ or $e \in E(S)$.

(1) If $e = \square$, we define $\beta((\square, s, f)) = \tau_\mathbf{P}(\gamma((\square, s, f)))$.

(2) If $e \in E(S)$, we define $\beta((e, s, f)) = \delta_\mathbf{P}(\lceil e \rceil^k, \gamma((e, s, f)))$.

It remains to show that for any well-formed word $w \in \mathbb{A}_\alpha^+$, we have $\tau_\mathbf{P}(\gamma(w)) = \beta(w)$. We have $w = b_1 \cdots b_n$ with $b_1, \ldots, b_n \in \mathbb{A}_\alpha$. We show that for any $\ell \leq n$, we have:

$$\tau_\mathbf{P}(\gamma(b_1 \cdots b_\ell)) = \beta(b_1 \cdots b_\ell).$$

We shall argue by induction on ℓ. The case $\ell = n$, will then yield the desired result. In the base case $\ell = 1$, since w is well-formed, we know that $b_1 = (\square, s, f)$ for some $s \in S^1$ and $f \in E(S) \cup \{\square\}$.

Thus, it is immediate by definition of β that we have $\tau_{\mathbf{P}}(\gamma(b_1)) = \beta(b_1)$. Assume now that $\ell \geq 2$. Since β is a morphism, we have

$$\beta(b_1 \cdots b_\ell) = \beta(b_1 \cdots b_{\ell-1}) \cdot \beta(b_\ell).$$

It then follows from the induction hypothesis that,

$$\beta(b_1 \cdots b_\ell) = \tau_{\mathbf{P}}(\gamma(b_1 \cdots b_{\ell-1})) \cdot \beta(b_\ell).$$

Since w is well-formed, we know that there exist $f, g \in E(S) \cup \{\square\}$, $s, t \in S^1$ and $e \in E(S)$ such that $b_{\ell-1} = (g, t, e)$ and $b_\ell = (e, s, f)$. Hence, $\beta(b_\ell) = \delta_{\mathbf{P}}(\lceil e \rceil^k, \gamma(b_\ell))$. Moreover, since $b_{\ell-1} = (g, t, e)$, it follows from the definition of γ that $\lceil e \rceil^k$ is a suffix of $\gamma(b_1 \cdots b_{\ell-1})$. Thus, since $\mathbf{P}$ is a SU_k-partition of A^* and $\lceil e \rceil^k$ has length at least k (this is where $\lceil e \rceil$ being nonempty is crucial), we have $[\lceil e \rceil^k]_{\mathbf{P}} = [\gamma(b_1 \cdots b_{\ell-1})]_{\mathbf{P}}$. Hence,

$$\beta(b_\ell) = \delta_{\mathbf{P}}(\lceil e \rceil^k, \gamma(b_\ell)) = \delta_{\mathbf{P}}(\gamma(b_1 \cdots b_{\ell-1}), \gamma(b_\ell)).$$

Altogether, this yields,

$$\beta(b_1 \cdots b_\ell) = \tau_{\mathbf{P}}(\gamma(b_1 \cdots b_{\ell-1})) \cdot \delta_{\mathbf{P}}(\gamma(b_1 \cdots b_{\ell-1}), \gamma(b_\ell)).$$

By Lemma 4.4, this says that $\beta(b_1 \cdots b_\ell) = \tau_{\mathbf{P}}(\gamma(b_1 \cdots b_\ell))$, which concludes the proof.

4.5 From C-covering to $(C \circ \mathbf{SU})$-covering

We now turn to the direction $(3) \Rightarrow (2)$ in Theorem 4.12. The argument is based on the following proposition which states a generic property of the morphism $\alpha : A^* \to M$.

PROPOSITION 4.18. *There exists a map $\eta : A^* \to \mathbb{A}_\alpha^*$ such that:*

(1) *For any $L \subseteq A^*$ recognized by α, we have $L = \eta^{-1}(\mathrm{WF}_\alpha(L))$.*

(2) *For any $K \in C(\mathbb{A}_\alpha)$, we have $\eta^{-1}(K) \in (C \circ \mathrm{SU}_{2|M|})(A)$.*

Before we prove Proposition 4.15, we use it to finish the proof of Theorem 4.12. Consider a language L and a multiset of languages $\mathbf{L}$, all recognized by α. Moreover, assume that $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$ is C-coverable. We have to show that $(L, \mathbf{L})$ is $(C \circ \mathrm{SU}_{2|M|})$-coverable.

By hypothesis, there exists a separating C-cover $\mathbf{K}_{\mathbb{A}_\alpha}$ of $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$, we use it to build a separating $(C \circ \mathrm{SU}_{2|M|})$-cover $\mathbf{K}_A$ of $(L, \mathbf{L})$. Consider the map $\eta : A^* \to \mathbb{A}_\alpha^*$ given by Proposition 4.18. We define,

$$\mathbf{K}_A = \{\eta^{-1}(K) \mid K \in \mathbf{K}_{\mathbb{A}_\alpha}\}.$$

To conclude the proof, we show that $\mathbf{K}_A$ is a separating $(C \circ \mathrm{SU}_{2|M|})$-cover of $(L, \mathbf{L})$.

Let us first prove that $\mathbf{K}_A$ is a $(C \circ \mathrm{SU}_{2|M|})$-cover of L. Let $w \in L$, we first have to find $H \in \mathbf{K}_A$ such that $w \in H$. Since $w \in L$, we know from the first item in Proposition 4.18 that $\eta(w) \in \mathrm{WF}_\alpha(L)$. Hence, since $\mathbf{K}_{\mathbb{A}_\alpha}$ is a cover of $\mathrm{WF}_\alpha(L)$, there exists $K \in \mathbf{K}_{\mathbb{A}_\alpha}$ such that $\eta(w) \in K$. It now follows that w belongs to $\eta^{-1}(K) \in \mathbf{K}_A$. We conclude that $\mathbf{K}_A$ is a cover of L. Moreover, it is a $(C \circ \mathrm{SU}_{2|M|})$-cover by Item (2) in Proposition 4.18.

We now prove that $\mathbf{K}_A$ is separating. Let $H \in \mathbf{K}_A$, we have to find $L' \in \mathbf{L}$ such that $L' \cap H = \emptyset$. By definition $H = \eta^{-1}(K)$ for some $K \in \mathbf{K}_{\mathbb{A}_\alpha}$. Since $\mathbf{K}_{\mathbb{A}_\alpha}$ is a separating for $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$, we know that there exists $L \in \mathbf{L}$ such that $K \cap \mathrm{WF}_\alpha(L) = \emptyset$. It is immediate that $\eta^{-1}(K) \cap L = \emptyset$ since $L = \eta^{-1}(\mathrm{WF}_\alpha(L))$ by Item (1) in Proposition 4.18.

It remains to prove Proposition 4.18, to which we devote the rest of this section.

Proof of Proposition 4.18: definition of η. We begin by defining the map $\eta : A^* \to \mathbb{A}_\alpha^*$. Let us point out that η is **not** be a morphism (otherwise, since C is closed under inverse image, all $\eta^{-1}(K)$ would belong to $C(A)$, which is not the case in general). We start with a preliminary definition.

Given a word w, a position x in w (*i.e.*, $x \in \{0, 1, \ldots, |w| - 1\}$) and a natural number $k \in \mathbb{N}$, we define the *k-type* of x as the following word of length at most k:

- If $x < k$, then the k-type of x is the prefix $w[0, x - 1]$ of length x.
- If $x \geq k$, then the k-type of x is the infix $w[x - k, x - 1]$ of length k.

For the construction of η, we fix $k = |M|$. Moreover, we choose an arbitrary order on the set of idempotents $E(S)$ (recall that $S = \alpha(A^+)$).

Consider a nonempty word $w \in A^+$ and a position x in w. We say that x is *distinguished* when there exists an idempotent $e \in E(S)$ such that the k-type u of x satisfies $\alpha(u) \cdot e = \alpha(u)$. The following fact states that distinguished positions are frequent.

FACT 4.19. *Let $w \in A^+$ be such that $|w| \geq k$ and let $y \geq k - 1$ be a position in w. Then, there exists a distinguished position x in w such that $y - (k - 1) \leq x \leq y$.*

PROOF. This follows from the pigeonhole principle. By definition, $w[y - (k - 1), y]$ is a word $a_1 \cdots a_k$ of length k. For all $1 \leq j \leq k$, we define $w_j = a_1 \cdots a_j$. Moreover, we let $w_0 = \varepsilon$. By definition, we have $k = |M|$. Thus, we obtain from the pigeonhole principle that there exist $0 \leq j_1 < j_2 \leq k$ such that, $\alpha(w_{j_1}) = \alpha(w_{j_2})$.

We claim that the position $x = y - (k - j_1 - 1)$ in w is distinguished. Indeed, by definition, we have $w_{j_2} = w_{j_1} a_{j_1+1} \cdots a_{j_2}$. Therefore,

$$
\begin{aligned}
\alpha(w_{j_1}) &= \alpha(w_{j_2}) \\
&= \alpha(w_{j_1}) \cdot \alpha(a_{j_1+1} \cdots a_{j_2}) \\
&= \alpha(w_{j_1}) \cdot (\alpha(a_{j_1+1} \cdots a_{j_2}))^p \quad \text{for all } p \in \mathbb{N}.
\end{aligned}
$$

It is standard (and easy to check) that there exists $p \geq 1$ such that $(\alpha(a_{j_1+1} \cdots a_{j_2}))^p$ is an idempotent $e \in E(S)$. Therefore, since w_{j_1} is a suffix of the k-type of the position $x = y - (k - j_1 - 1)$, we know that x is distinguished, as witnessed by the idempotent e. $\square$

We are now ready to define the map $\eta : A^* \to \mathbb{A}_\alpha^*$. Consider a word $w \in A^*$. If w does not contain any distinguished position, we define,

$$
\eta(w) = (\square, \alpha(w), \square).
$$

Otherwise, w has $n \geq 1$ distinguished positions, say $x_0 < \cdots < x_{n-1}$. We let $u_0, \cdots, u_{n-1}$ be their respective k-types. Finally, let $e_0, \ldots, e_{n-1} \in E(S)$ be such that for all $i \geq 0$, e_i is the smallest idempotent (according to the arbitrary order that we fixed over $E(S)$) such that $\alpha(u_i) \cdot e_i = \alpha(u_i)$. We define $\eta(w) \in \mathbb{A}_\alpha^+$ as the following well-formed word:

$$
\eta(w) = (\square, \alpha(w_0), e_0) \cdot (e_0, \alpha(w_1), e_1) \cdots (e_{n-2}, \alpha(w_{n-1}), e_{n-1}) \cdot (e_{n-1}, \alpha(w_n), \square) \tag{3}
$$

where $w_0, \ldots, w_n$ are the unique words such that w may be decomposed as $w = w_0 w_1 \cdots w_n$ where for all $i \geq 0$, the word w_{i+1} starts at position x_i in w. In other terms, $w_0 = w[0, x_0 - 1]$, $w_i = w[x_{i-1}, x_i - 1]$ for $1 \leq i \leq n - 1$ and $w_n = w[x_{n-1}, |w| - 1]$. Observe that for any $w \in A^*$, $\eta(w) \in \mathbb{A}_\alpha^+$ is well-formed by construction.

It remains to prove that this definition satisfies the two items in Proposition 4.18.

Proof of Proposition 4.18: first item. Consider a language $L \subseteq A^*$ recognized by α, we have to show that $L = \eta^{-1}(\text{WF}_\alpha(L))$. This amounts to proving that for any $w \in A^*$, $w \in L$ if and only if $\eta(w) \in \text{WF}_\alpha(L)$.

Since α recognizes L, $w \in L$ if and only if $\alpha(w) \in \alpha(L)$. Moreover, since $\eta(w)$ is well-formed, by definition, $\eta(w) \in \text{WF}_\alpha(L)$ if and only if $\text{EVAL}(\eta(w)) \in \alpha(L)$. Hence, it suffices to prove that $\text{EVAL}(\eta(w)) = \alpha(w)$. This is immediate from the definition if w has no distinguished position. Otherwise, w may be decomposed as $w = w_0 \cdots w_n$ and (3) holds. By choice of the idempotents

used in the construction, we have $\alpha(w_1 \cdots w_i) \cdot e_i = \alpha(w_1 \cdots w_i)$ for all $i \geq 0$. Hence, it is immediate from a simple induction that

$$\text{EVAL}(\eta(w)) = \alpha(w_0) \cdot e_0 \cdot \alpha(w_1) \cdot e_1 \cdot \alpha(w_2) \cdot \cdots \cdot e_{n-1} \cdot \alpha(w_n) = \alpha(w_1 \cdots w_n) = \alpha(w).$$

Proof of Proposition 4.18: second item. Given an arbitrary language $K \in C(\mathbb{A}_\alpha)$, we have to prove that $\eta^{-1}(K)$ belongs to $(C \circ \text{SU}_{2k})(A)$ (recall that we fixed $k = |M|$).

By definition of $C \circ \text{SU}_{2k}$, the first thing we have to do is choose some SU_{2k}-partition of A^*. Recall that $\sim_{2k}$ denotes the canonical equivalence associated to SU_{2k}: given $w, w' \in A^*$, we have $w \sim_{2k} w'$ when $w \in L \Leftrightarrow w' \in L$ for any $L \in \text{SU}_{2k}$. We denote by $\mathbf{P}$ the partition of A^* into $\sim_{2k}$-classes. By Lemma 2.10, $\mathbf{P}$ is a SU_{2k}-partition of A^*. It now remains to exhibit languages $L_P \in C(\mathbf{P} \times A)$ for all $P \in \mathbf{P}$ such that

$$\eta^{-1}(K) = \bigcup_{P \in \mathbf{P}} \left(P \cap \tau_{\mathbf{P}}^{-1}(L_P) \right).$$

We start with preliminary definitions. We know that for any word $w \in A^*$ having n distinguished positions, $\eta(w) \in \mathbb{A}_\alpha^+$ has length $n+1$ and is built by decomposing w according to these distinguished positions. We let $\eta_p(w) \in \mathbb{A}_\alpha^*$ be the (possibly empty) prefix of $\eta(w)$ made of the first n letters of $\eta(w)$ ($\eta_p(w)$ is not well-formed: the third component of the rightmost letter is not "$\square$"). The argument is now based on the two following lemmas.

LEMMA 4.20. *Let $P \in \mathbf{P}$. Then there exists a letter $b_P \in \mathbb{A}_\alpha$ such that for any $w \in P$, the rightmost letter in $\eta(w)$ is b_P, i.e., $\eta(w) = \eta_p(w) \cdot b_P$.*

LEMMA 4.21. *There exists a morphism $\beta : (\mathbf{P} \times A)^* \to \mathbb{A}_\alpha^*$ such that for any $w \in A^*$, $\beta(\tau_{\mathbf{P}}(w)) = \eta_p(w)$.*

Before proving these lemmas, let us use them to finish the proof that $\eta^{-1}(K)$ belongs to $(C \circ \text{SU}_{2k})(A)$. For any $P \in \mathbf{P}$, we let $b_P \in \mathbb{A}_\alpha$ be as defined in Lemma 4.20. Moreover, let $\beta : (\mathbf{P} \times A)^* \to \mathbb{A}_\alpha^*$ be the morphism described in Lemma 4.21. We claim that:

$$\eta^{-1}(K) = \bigcup_{P \in \mathbf{P}} \left(P \cap \tau_{\mathbf{P}}^{-1}(\beta^{-1}(K(b_P)^{-1})) \right).$$

This concludes the proof: since C is closed under right quotient and inverse image, we know that for any $P \in \mathbf{P}$, $\beta^{-1}(K(b_P)^{-1}) \in C(\mathbf{P} \times A)$. Thus, it is immediate that $\eta^{-1}(K)$ belongs to $(C \circ \text{SU}_{2k})(A)$ by definition.

Let us prove the claim. Consider a word $w \in A^*$ and let P be the unique language in the partition $\mathbf{P}$ of A^* such that $w \in P$. It suffices to show that $w \in \eta^{-1}(K)$ if and only if $w \in \tau_{\mathbf{P}}^{-1}(\beta^{-1}(K(b_P)^{-1}))$. By Lemma 4.20, we know that $w \in \eta^{-1}(K)$ if and only if $\eta_p(w) \cdot b_P \in K$, i.e., $\eta_p(w) \in K(b_P)^{-1}$. Finally, since $\beta(\tau_{\mathbf{P}}(w)) = \eta_p(w)$ by Lemma 4.21, this is equivalent to $w \in \tau_{\mathbf{P}}^{-1}(\beta^{-1}(K(b_P)^{-1}))$, which concludes the proof.

It remains to prove Lemmas 4.20 and 4.21.

Proof of Lemma 4.20. Since any $P \in \mathbf{P}$ is by definition a $\sim_{2k}$-class, this amounts to proving that given $w, w' \in A^*$ such that $w \sim_{2k} w'$, $\eta(w)$ and $\eta(w')$ have the same rightmost letter. We consider two possible cases.

If w has no distinguished position, then we have $|w| < k$ by Fact 4.19, hence $\{w\} \in \text{SU}_{2k}$. Since $w \sim_{2k} w'$, we have $w' \in \{w\}$, i.e., $w = w'$. The result is now immediate.

Assume on the contrary that w contains at least one distinguished position. We let x be the rightmost one, and u be the k-type of x. By definition, the rightmost letter in $\eta(w)$ is $(e, \alpha(v), \square)$ where $v = w[x, |w| - 1]$ and $e \in E(S)$ the smallest idempotent such that $\alpha(u) \cdot e = \alpha(u)$. Note that uv is a suffix of w by definition. Since x is the rightmost distinguished position by definition,

it follows from Fact 4.19 that $|v| \leq k$ (otherwise, there would be another distinguished position strictly to the right of x). It follows that $|uv| \leq 2k$. Thus, since $w \sim_{2k} w'$, uv is a suffix of w' as well. It now follows from the definitions that the rightmost letter in w' must be $(e, \alpha(v), \square)$ as well. This concludes the proof of Lemma 4.20.

Proof of Lemma 4.21. Let us start with a few simple observations. Consider some word $w \in A^*$. By definition, if w has no distinguished position, then $\eta_p(w) = \varepsilon$. Otherwise, w has $n \geq 1$ distinguished positions $x_0 < \cdots < x_{n-1}$ and,

$$\eta_p(w) = (\square, \alpha(w_0), e_0) \cdot (e_0, \alpha(w_1), e_1) \cdots (e_{n-2}, \alpha(w_{n-1}), e_{n-1})$$

where $w_0 = w[0, x_0 - 1]$, $w_i = w[x_{i-1}, x_i - 1]$ for $1 \leq i \leq n - 1$. Note that there is a natural bijection between the distinguished positions of w and the positions of $\eta_p(w)$, which associates to any distinguished position x_i in w the position $\lfloor x_i \rfloor = i$ in $\eta_p(w)$.

By definition any position x in w may also be viewed as a position of $\tau_{\mathbf{P}}(w) \in (\mathbf{P} \times A)^*$. Because of our choice of k as $|M|$, the map $w \mapsto \eta_p(w)$ is designed so that for any position x in w, whether x is distinguished and if so the label of $\lfloor x \rfloor$ in $\eta_p(w)$ depends only on the label of x in $\tau_{\mathbf{P}}(w)$. Let us state this property in the following lemma.

LEMMA 4.22. *For any letter $(P, a) \in \mathbf{P} \times A$, one of the two following properties hold:*
(1) For any $w \in A^*$ and any position x in w, if x has label (P, a) in $\tau_{\mathbf{P}}(w)$, then x is not distinguished.
(2) There is a letter $c_{(P,a)} \in \mathbb{A}_\alpha$ such that for any $w \in A^*$ and any position x in w, if x has label (P, a) in $\tau_{\mathbf{P}}(w)$, then x is distinguished and $\lfloor x \rfloor$ has label $c_{(P,a)}$ in $\eta_p(w)$.

Before proving Lemma 4.22, we first use it to define the morphism $\beta : (\mathbf{P} \times A)^* \to \mathbb{A}_\alpha^*$ and finish the argument for Lemma 4.21. We have to define the image of each letter in $\mathbf{P} \times A$. Let $(P, a) \in \mathbf{P} \times A$ be a letter.

- If (P, a) satisfies the first item Lemma 4.22, we let $\beta((P, a)) = \varepsilon$.
- If (P, a) satisfies the first second item in Lemma 4.22, we let $\beta((P, a)) = c_{(P,a)}$.

It is now immediate from Lemma 4.22 that β satisfies the desired property: for any $w \in A^*$, $\beta(\tau_{\mathbf{P}}(w)) = \eta_p(w)$.

It remains to prove Lemma 4.22. By definition of the partition $\mathbf{P}$, this amounts to proving that given $w, w' \in A^*$ and x, x' positions in w, w' such that $w[0, x-1] \sim_{2k} w'[0, x'-1]$, x is distinguished if and only if x' is distinguished and in that case, $\lfloor x \rfloor$ and $\lfloor x' \rfloor$ in $\eta_p(w)$ and $\eta_p(w')$ have the same label. If $w[0, x - 1] \sim_{2k} w'[0, x' - 1]$, then x and x' have the same k-type, and since the k-type of a position determines whether it is distinguished or not, x and x' are either both distinguished, or none of them is.

We now concentrate on the second property. Assume that x and x' are distinguished, we show that the positions $\lfloor x \rfloor$ and $\lfloor x' \rfloor$ in $\eta_p(w)$ and $\eta_p(w')$ carry the same label. We define u as the common k-type of x and x' and $e \in E(S)$ as the smallest idempotent such that $\alpha(u) \cdot e = \alpha(u)$. We distinguish two cases.

Assume first that x is the leftmost distinguished position in w. It follows that the label of $\lfloor x \rfloor$ is $(\square, \alpha(w[0, x - 1]), e)$. We have to show that this is also the label of $\lfloor x' \rfloor$. By hypothesis, we know that the prefix $w[0, x - 1]$ contains no distinguished position. Thus, Fact 4.19 yields that $w[0, x-1] < k$. Since $w[0, x-1] \sim_{2k} w'[0, x'-1]$, it is then immediate that $w[0, x-1] = w'[0, x'-1]$. We conclude that x' is also the leftmost distinguished position of w' and that $\lfloor x \rfloor, \lfloor x' \rfloor$ have the same label, namely $(\square, \alpha(w[0, x - 1]), e)$.

Assume now that x is not the leftmost distinguished position in w. Let y be the distinguished position which directly precedes x in w. Furthermore, let v be the k-type of y and $f \in E(S)$ be the smallest idempotent such that $\alpha(v) \cdot f = \alpha(v)$. By definition, the label of $\lfloor x \rfloor$ is $(f, \alpha(w[y, x - 1]), e)$.

We have to show that this is also the label of $\lfloor x' \rfloor$. By Fact 4.19, we have $|w[y, x-1]| \leq k$ (otherwise, there would be a third distinguished position in w strictly between y and x, contradicting the definition of y). Therefore, $|v \cdot w[y, x-1]| \leq 2k$. Moreover, $v \cdot w[y, x-1]$ is a suffix of $w[0, x-1]$ by definition. Since $w[0, x-1] \sim_{2k} w'[0, x'-1]$, we obtain that $v \cdot w[y, x-1]$ is also a suffix of $w'[0, x'-1]$. This shows that $\lfloor x' \rfloor$ has label $(f, \alpha(w[y, x-1]), e)$ as well, which concludes the proof.

5 APPLICATION TO TWO-VARIABLE FIRST-ORDER LOGIC

This is the first of two sections in which we illustrate Theorem 4.12 and use it to obtain algorithms for a particular class of languages. Here, we consider the two-variable fragment of first-order logic over words (defined in Section 3). Specifically, we show that the covering is decidable for the strong variant: $FO^2(<, +1)$. Let us state this result.

COROLLARY 5.1 (OF THEOREM 4.12). *Covering and separation are decidable for* $FO^2(<, +1)$ *over words.*

As we explained in Remark 4.14, using Theorem 4.12 to obtain Corollary 5.1 requires clearing two preliminary steps. First we need to that $FO^2(<, +1)$ is the SU-enrichment of some lattice closed under right quotient and inverse image (namely $FO^2(<)$ in this case). Then, we need to show that covering and separation are decidable for $FO^2(<)$.

Fortunately, the second step has already been achieved: it was shown in [28] and in [32, 34] that $FO^2(<)$-separation and $FO^2(<)$-covering are decidable. Thus, we just have to show that $FO^2(<, +1)$ is the SU-enrichment of $FO^2(<)$. We state this in the following proposition.

PROPOSITION 5.2. *Over words,* $FO^2(<, +1)$ *is the SU-enrichment of* $FO^2(<)$.

REMARK 5.3. *It is important to point out that while the formulation is new, the underlying ideas behind Proposition 5.2 were already known. This connection between* $FO^2(<)$ *and* $FO^2(<, +1)$ *was originally presented by Thérien and Wilke [44]. However, the full proof of this result is scattered in the literature and relies on different terminology. Thus, it makes sense to detail it here.*

There are two inclusions to prove for showing Proposition 5.2. We devote a subsection to each of them.

5.1 From enrichment to successor

We show here that any language in the SU-enrichment of $FO^2(<)$ may be defined by an $FO^2(<, +1)$ sentence. For this, let us fix an alphabet A and consider a language $L \in FO^2(<) \circ SU$ over A. By definition, there exists an SU-partition $\mathbf{P}$ of A^* such that,

$$L = \bigcup_{P \in \mathbf{P}} (P \cap \tau_{\mathbf{P}}^{-1}(L_P))$$

where all languages $L_P \subseteq (\mathbf{P} \times A)^*$ are definable in $FO^2(<)$. We show that L can be defined by an $FO^2(<, +1)$ sentence. Since we may freely use Boolean connectives in $FO^2(<, +1)$ sentences, it suffices to show that for all $P \in \mathbf{P}$, both P and $\tau_{\mathbf{P}}^{-1}(L_P)$ are defined by an $FO^2(<, +1)$ sentence. We start with the following preliminary lemma.

LEMMA 5.4. *For any* $u \in A^*$, *one may construct an* $FO^2(<, +1)$ *formula* $\varphi_u(x)$ *(with one free variable* x*) such that for any* $w \in A^*$ *and any position* x *in* u, *we have* $w \models \varphi_u(x)$ *if and only if* $w[0, x-1] \in A^*u$.

PROOF. We use induction on u to define $\varphi_u(x)$. If $u = \varepsilon$, it suffices to define $\varphi_u(x) = \top$. Otherwise, $u = va$ for some $v \in A^*$ and $a \in A$ and we define,

$$\varphi_u(x) = \exists y \, (y + 1 = x \wedge a(y) \wedge \varphi_v(y)).$$

This concludes the proof of Lemma 5.4.   □

We now start the main argument. Let $P \in \mathbf{P}$, we first show that P and $\tau_{\mathbf{P}}^{-1}(L_P)$ may both be defined by an $\mathrm{FO}^2(<, +1)$ sentence.

Case 1: each $P \in \mathbf{P}$ may be defined by an $\mathrm{FO}^2(<, +1)$ sentence. By definition of $\mathbf{P}$, we know that $P \in \mathrm{SU}(A)$, whence P is a finite Boolean combination of languages of the form $A^* w$, with $w \in A^*$. Since $\mathrm{FO}^2(<, +1)$ is a Boolean algebra, it suffices to show that $A^* w$ can be defined in $\mathrm{FO}^2(<, +1)$. If $w = \varepsilon$, then A^* is defined by the sentence $\top$. Otherwise, $w = ua$ with $u \in A^*$ and $a \in A$, $A^* w$ is defined by $\exists x \; max(x) \wedge a(x) \wedge \varphi_u(x)$.

Case 2: the language $\tau_{\mathbf{P}}^{-1}(L_P)$ may be defined by an $\mathrm{FO}^2(<, +1)$ sentence. Recall that $L_P \subseteq (\mathbf{P} \times A)^*$ is defined by some $\mathrm{FO}^2(<)$ sentence ξ. We use the following fact which is an immediate consequence of Lemma 5.4 since all languages in $\mathbf{P}$ belong to SU.

FACT 5.5. *Given any $(P, a) \in \mathbf{P} \times A$, there exists a $\mathrm{FO}^2(<, +1)$ formula $\zeta_{(P,a)}(x)$ (over A) with one free variable such that for any $w \in A^*$ and any position x in w, we have, $w \models \zeta_{(P,a)}(x)$ if and only if x has label $(P, a) \in \tau_{\mathbf{P}}(w)$.*

It is now simple to construct an $\mathrm{FO}^2(<, +1)$ sentence defining $\tau_{\mathbf{P}}^{-1}(L_P)$ from the $\mathrm{FO}^2(<)$ sentence ξ defining $L_P \subseteq (\mathbf{P} \times A)$: we replace atomic subformulas of the form $(P, a)(x)$, for some $(P, a) \in \mathbf{P} \times A$, by the formula $\zeta_{(P,a)}(x)$. This concludes the proof for this direction.

5.2 Ehrenfeucht-Fraïssé games

To prove the converse direction in Proposition 5.2, we need the Ehrenfeucht-Fraïssé games associated to $\mathrm{FO}^2(<)$ and $\mathrm{FO}^2(<, +1)$. We first define these games.

REMARK 5.6. *For the sake of simplifying the $\mathrm{FO}^2(<, +1)$-game, we shall assume that the predicates min, max and ε are not allowed in $\mathrm{FO}^2(<, +1)$. This is not restrictive since $min(x)$ is defined by $\neg(\exists y \; y < x)$, $max(x)$ by $\neg(\exists y \; y > x)$ and ε by $\forall x \bot$.*

The *(quantifier) rank* of a first-order formula φ, denoted $\mathrm{rank}(\varphi)$, is defined as the largest number of quantifiers along a branch in the parse tree of φ. Formally, $\mathrm{rank}(\varphi) = 0$ if φ is an atomic formula, $\mathrm{rank}(\neg \varphi) = \mathrm{rank}(\varphi)$, $\mathrm{rank}(\varphi_1 \vee \varphi_2) = \max(\mathrm{rank}(\varphi_1), \mathrm{rank}(\varphi_2))$ and $\mathrm{rank}(\exists x \; \varphi) = \mathrm{rank}(\varphi) + 1$. For any alphabet A, any natural number $k \in \mathbb{N}$ and any two words $w, w' \in A^*$, we write:

- $w \cong_k w'$ when w and w' satisfy the same $\mathrm{FO}^2(<)$ sentences of rank k.
- $w \cong_k^+ w'$ when w and w' satisfy the same $\mathrm{FO}^2(<, +1)$ sentences of rank k.

It is immediate that both $\cong_k$ and $\cong_k^+$ are equivalence relations over the set A^*. Moreover, one may verify the following standard lemma, which characterizes languages definable in $\mathrm{FO}^2(<)$ and $\mathrm{FO}^2(<, +1)$ using these relations:

LEMMA 5.7 (FOLKLORE). *Given any alphabet A, any language $L \subseteq A^*$ and any natural number $k \in \mathbb{N}$, the following properties hold:*

- *L may be defined by a $\mathrm{FO}^2(<)$ sentence of rank k if and only if L is a union of $\cong_k$-classes.*
- *L may be defined by a $\mathrm{FO}^2(<, +1)$ sentence of rank k if and only if L is a union of $\cong_k^+$-classes.*

We now define the Ehrenfeucht-Fraïssé games associated to $\mathrm{FO}^2(<)$ and $\mathrm{FO}^2(<, +1)$, which give alternate definitions for the relations $\cong_k$ and $\cong_k^+$.

$\mathrm{FO}^2(<)$ game. The board of the $\mathrm{FO}^2(<)$-game consists of two words w and w'. It lasts a predefined number k of rounds. There are two players called *Spoiler and Duplicator*. Moreover, there are two pebbles and at any time during the game after the first round, one of them is placed on a position of w and the other on a position w', and these two positions have the same label (when the game starts, no pebble is on the board).

In the first round, Spoiler chooses a word (either w or w') and places a pebble on a position of this word. Duplicator must answer by placing the other pebble on a position of the other word having the same label. The remaining rounds are played as follows. Spoiler chooses a word (either w or w') and moves the pebble inside this 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' having the same label as y and such that $x' < y'$ if and only if $x < y$.

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

FO^2($<$, +1) game. The FO^2($<$, +1)-game is defined similarly with an additional constraint for Duplicator when answering Spoiler's moves. When Spoiler makes a move, Duplicator must choose her answer y' so that $x' + 1 = y'$ if and only if $x + 1 = y$ and $y' + 1 = x'$ if and only if $y + 1 = x$ (in addition to the constraints already presented for the FO^2($<$) game).

We may now state the Ehrenfeucht-Fraïssé theorem for FO^2($<$) and FO^2($<$, +1): $\cong_k$ and $\cong_k^+$ are characterized by the FO^2($<$) game and the FO^2($<$, +1) game, respectively.

THEOREM 5.8 (FOLKLORE). *Let A be an alphabet. Given any $k \in \mathbb{N}$ and any words $w, w' \in A^*$, the two following properties hold:*

- $w \cong_k w'$ *if and only if Duplicator has a winning strategy for playing k-rounds in the* FO^2($<$)-*game over w and w'.*
- $w \cong_k^+ w'$ *if and only if Duplicator has a winning strategy for playing k-rounds in the* FO^2($<$, +1)-*game over w and w'.*

5.3 From successor to enrichment

We are now ready to show the remaining direction in Proposition 5.2: any language that can be defined by an FO^2($<$, +1) sentence belongs to the SU-enrichment of FO^2($<$).

We start with a preliminary definition. Recall that given any $k \in \mathbb{N}$, $\sim_k$ denotes the canonical equivalence defined on A^* associated to SU$_k$. We write $\mathbf{P}_k$ for the finite partition of A^* into classes of $\sim_k$. Note that $\mathbf{P}_k$ is a SU-partition of A^* by Lemma 2.10. In the proof, we use the SU-partitions $\mathbf{P}_k$ to build languages in FO^2($<$) ∘ SU. For the sake of simplifying the notation, given $k \in \mathbb{N}$, we shall write τ_k for the map $\tau_{\mathbf{P}_k} : A^* \to (\mathbf{P}_k \times A^*)$.

Our argument to prove FO^2($<$, +1) $\subseteq$ FO^2($<$) ∘ SU is based on the following result.

PROPOSITION 5.9. *Let $k \in \mathbb{N}$, and $w, w' \in A^*$. If $\tau_{2k}(w) \cong_k \tau_{2k}(w')$, then $w \cong_k^+ w'$.*

Before showing Proposition 5.9, we use it to conclude our argument for Proposition 5.2. Let $L \subseteq A^*$ be defined by some FO^2($<$, +1) sentence φ. We show that $L \in$ FO^2($<$) ∘ SU. By Lemma 5.7, L is a union of $\cong_k^+$-classes, where k is the rank of φ. We define a language $H \subseteq (\mathbf{P}_{2k} \times A)^*$ as follows:

$$H = \{u \in (\mathbf{P}_{2k} \times A)^* \mid \text{there exists } w \in L \text{ such that } \tau_{2k}(w) \cong_k u\}.$$

By definition, H is a union of $\cong_k$-classes and can therefore by defined by some FO^2($<$) sentence of rank k (see Lemma 5.7). We show that,

$$L = \tau_{2k}^{-1}(H) = \bigcup_{P \in \mathbf{P}_{2k}} (P \cap \tau_{2k}^{-1}(H)), \tag{4}$$

an expression showing that $L \in$ FO^2($<$) ∘ SU (since $\mathbf{P}_{2k}$ is an SU-partition of A^*). To prove (4), we start with the left to right inclusion. Assume that $v \in L$. It is then immediate from the definition that $\tau_{2k}(v) \in H$ since $\tau_{2k}(v) \cong_k \tau_{2k}(v)$, hence we get $v \in \tau_{2k}^{-1}(H)$, which establishes the first inclusion. For the converse inclusion, assume that $v \in \tau_{2k}^{-1}(H)$. We show that $v \in L$. Since $v \in \tau_{2k}^{-1}(H)$, we

get by definition of H that there exists $w \in L$ such that $\tau_{2k}(w) \cong_k \tau_{2k}(v)$. Thus, we obtain from Proposition 5.9 that $w \cong_k^+ v$. Finally, since $w \in L$ and L is a union of $\cong_k^+$-classes, we obtain that $v \in L$.

It remains to prove Proposition 5.9. Let $k \in \mathbb{N}$ and let $w, w' \in A^*$ be words such that $\tau_{2k}(w) \cong_k \tau_{2k}(w')$. Our objective is to prove that $w \cong_k^+ w'$. By Theorem 5.8, this amounts to describing a winning strategy for Duplicator in the k-round $\mathrm{FO}^2(<, +1)$-game over w and w'. We call this game $\mathcal{G}$. Duplicator's strategy involves playing another "shadow" $\mathrm{FO}^2(<)$-game over $\tau_{2k}(w)$ and $\tau_{2k}(w')$. Recall that by hypothesis and Theorem 5.8, she has a winning strategy for k rounds in this shadow game. Depending on the moves that Spoiler makes in $\mathcal{G}$, Duplicator may have to simulate a move by Spoiler in the shadow game. Her strategy in the shadow game gives her an answer to this simulated move, which she is then able to use for computing a suitable answer in $\mathcal{G}$.

Recall that for any word $v \in A^*$ (including w and w'), $\tau_{2k}(u)$ is a relabeling of u over $(\mathbf{P} \times A)$. In particular, this means that any position of w (resp. w') corresponds to a position of $\tau_{2k}(w)$ (resp. $\tau_{2k}(w')$), and may be viewed as such, and *vice versa*. This will be convenient to relate the moves performed in $\mathcal{G}$ to the ones played in the shadow game.

Given two positions x and z of w and a natural number $h \leq k - 1$, we say that *x is h-safe for z* when all positions y in u such that $|x - y| \leq h$ satisfy $z - 2k \leq y \leq z$. Note that these positions in w are all fully described by (the two components of) the label of z in $\tau_{2k}(w)$. We extend the definition to positions x', z' of w' in the same way.

We may now describe Duplicator's strategy in $\mathcal{G}$. It involves enforcing an invariant $\mathcal{I}(j)$ which has to hold after each round j. Let $1 \leq j \leq k$. Assume that j rounds have been played in $\mathcal{G}$ so far and let x, x' be the positions of w, w' on which the pebbles are currently placed in $\mathcal{G}$. Furthermore, let z, z' be the positions of $\tau_{2k}(w), \tau_{2k}(w')$ on which the pebbles are currently placed in the shadow game. We say that $\mathcal{I}(j)$ holds when the following conditions are met.

(1) $x - z = x' - z'$.

(2) x is $(k - j)$-safe for z and x' is $(k - j)$-safe for z'.

(3) Duplicator has a wining strategy for playing $k - j$ more rounds in the shadow game.

We now describe a strategy allowing Duplicator to play and enforce $\mathcal{I}(j)$ after each round j, $1 \leq j \leq k$. Let us first explain how Duplicator may enforce $\mathcal{I}(1)$ after round 1.

Assume that Spoiler puts a pebble on a position x of w (the case when Spoiler puts a pebble in w' is symmetrical). Then, Duplicator simulates a moves by Spoiler in the shadow game by putting a pebble on position $z = min(|w| - 1, x + k - 1)$ in $\tau_{2k}(w)$. She then obtains an answer z' in $\tau_{2k}(w')$ having the same label as z from her strategy. By definition, x is $(k - 1)$-safe for z. Since z and z' have the same label in $\tau_{2k}(w)$ and $\tau_{2k}(w')$, by definition of the labels in $\tau_{2k}(w)$ and $\tau_{2k}(w')$, there must exist a position x' in w' such that $z - x = z' - x'$ and with the same label as x. This position x' is Duplicator's answer, which is clearly correct. Moreover, $\mathcal{I}(1)$ is satisfied.

We now assume that $j \geq 1$ rounds have already been played and that $\mathcal{I}(j)$ holds. Let x, x' be the positions of w, w' on which the pebbles are currently placed in $\mathcal{G}$ and let z, z' be those in $\tau_{2k}(w), \tau_{2k}(w')$ on which the pebbles are currently placed in the shadow game. Assume that Spoiler moves the pebble from position x in w to a new position y (as before, the other case is symmetrical). We describe a correct answer for Duplicator which satisfies $\mathcal{I}(j + 1)$. There are two cases depending on whether y is $(k - j - 1)$-safe for z or not.

Assume first that y is $(k - j - 1)$-safe for z. In that case, Duplicator does not use the shadow game (the pebbles remain on z and z' for this round). Since z and z' have the same label in $\tau_{2k}(w)$ and $\tau_{2k}(w')$, there must exist a position y' in w' such that $z - y = z' - y'$ and with the same label

as y. This position y' is Duplicator's answer. One may verify that it is correct and that $\mathcal{I}(j+1)$ is satisfied.

We now assume that y is not $(k-j-1)$-safe for z. There are two sub-cases depending on whether $y < x$ or $x < y$. By symmetry, we only consider the case $x < y$. Note that by hypothesis, we have the following properties:

a) $y > x + 1$ (since x is $(k-j)$-safe, $x+1$ is $(k-j-1)$-safe).
b) $y > z - k + j + 1$ (as y is strictly right of the rightmost $(k-j-1)$-safe position for z).
c) $z \neq |w| - 1$ (otherwise, $y > x$ would be $(k-j-1)$-safe for z)

Duplicator first simulates a move by Spoiler in the shadow game: she moves the pebble in $\tau_{2k}(w)$ from z to $z_1 = min(|w|-1, y+k-j-1)$, i.e., to the leftmost position for which y is $(k-j-1)$-safe. Note that $z_1 > z$. Indeed, either $z_1 = |w|-1$ and $z < z_1$ by Item c) above. Otherwise, $z_1 = y+k-j-1 > z$ by Item b). Hence, Duplicator's strategy in the shadow game yields an answer $z_1' > z'$ in $\tau_{2k}(w')$ with the same label as z_1.

Since z_1, z_1' have the same label in $\tau_{2k}(w), \tau_{2k}(w')$, we now obtain a position y' in w' such that $z_1 - y = z_1' - y' = k - j - 1$ and with the same label as y. This position y' is Duplicator's answer. Proving that it is correct and that $\mathcal{I}(j+1)$ now holds amounts to showing that $y' > x' + 1$. Since x' was $(k-j)$-safe for z', we know that $x' \leq z' - k + j$. Since $z' < z_1'$, this yields $x' < z_1' - k + j$. Finally, we have $z_1' - y' = k - j - 1$ by definition. Altogether, this means that $x' + 1 < y'$, as desired.

This concludes the proof of Proposition 5.9, and therefore of Proposition 5.2 as well.

6 APPLICATION TO QUANTIFIER ALTERNATION

In this section, we illustrate Theorem 4.12 with a second example: the quantifier alternation of first-order logic over words. We prove the following result.

COROLLARY 6.1 (OF THEOREM 4.12). *Over words, covering and separation are decidable for the levels* $\Sigma_1(<, +1)$, $\mathcal{B}\Sigma_1(<, +1)$, $\Sigma_2(<, +1)$, $\mathcal{B}\Sigma_2(<, +1)$ *and* $\Sigma_3(<, +1)$ *in the alternation hierarchy of first-order logic.*

Note that Corollary 6.1 subsumes many difficult results from the literature. In particular, it shows that *membership* is decidable for $\mathcal{B}\Sigma_1(<, +1)$ and for $\Sigma_2(<, +1)$. Direct proofs for both of these results are difficult [12–14].

As usual, obtaining Corollary 6.1 from Theorem 4.12 requires clearing two preliminary steps.

- First we prove that the fragments mentioned in the theorem are the SU-enrichment of lattices closed under right quotient and inverse image. As expected, we use $\Sigma_1(<)$, $\mathcal{B}\Sigma_1(<)$, $\Sigma_2(<)$, $\mathcal{B}\Sigma_2(<)$ and $\Sigma_3(<)$ in this case.
- Then, covering and separation are shown to be decidable for these simpler classes.

Again, the second step has already been achieved: it is known that covering and separation are decidable for $\Sigma_1(<)$ [9, 36], $\mathcal{B}\Sigma_1(<)$ [9, 28, 36], $\Sigma_2(<)$ [29, 35, 36], $\mathcal{B}\Sigma_2(<)$ [36] and $\Sigma_3(<)$ [26, 27]. Thus, we concentrate on proving the connections with SU-enrichment. We state them in the following proposition.

PROPOSITION 6.2. *Given any* $n \geq 1$, *the following two properties hold over words:*

- $\Sigma_n(<, +1)$ *is the* SU*-enrichment of* $\Sigma_n(<)$.
- $\mathcal{B}\Sigma_n(<, +1)$ *is the* SU*-enrichment of* $\mathcal{B}\Sigma_n(<)$.

REMARK 6.3. *As for two-variable first-order logic in the previous section, these properties are essentially already known. The underlying ideas behind the connection with* SU*-enrichment are due to Straubing [42].*

In the rest of this section, we prove Proposition 6.2. We focus on the first item: $\Sigma_n(<, +1)$ is the SU-enrichment of $\Sigma_n(<)$. The proof for the second item is similar and left to the reader. There are two inclusions to prove, we devote one subsection to each of them.

6.1 From enrichment to successor

Let $n \in \mathbb{N}$ and consider some alphabet A. Let $L \subseteq A^*$ be a language belonging to $\Sigma_n(<) \circ \text{SU}$. We want to show that L is definable in $\Sigma_n(<, +1)$. By definition, there exists an SU-partition $\mathbf{P}$ of A^* such that

$$L = \bigcup_{P \in \mathbf{P}} (P \cap \tau_{\mathbf{P}}^{-1}(L_P)),$$

where all languages $L_P \subseteq (\mathbf{P} \times A)^*$ are definable in $\Sigma_n(<)$. We show that L can be defined by a $\Sigma_n(<, +1)$ sentence. Since we may freely use disjunction and conjunction in $\Sigma_n(<, +1)$ sentences, it suffices to show that for all $P \in \mathbf{P}$, both P and $\tau_{\mathbf{P}}^{-1}(L_P)$ are defined by a $\Sigma_n(<, +1)$ sentence.

Case 1. We start with the language $P \in \mathbf{P}$. By definition $P \in \mathbf{P}(A)$. It follows from Lemma 2.11 that P is a finite union of languages $\{w\}$ or A^*w where $w \in A^*$. Since $\Sigma_n(<, +1)$ is closed under union, it suffices to show that these two kinds of languages may be defined in $\Sigma_n(<, +1)$, which is easy: if $w = \varepsilon$, then $\{\varepsilon\}$ and A^* are defined by the sentences "ε" and "$\top$" respectively. Otherwise, $w = a_1 \cdots a_\ell$ for $a_1, \ldots, a_\ell \in A$. In that case, $\{w\}$ is defined by the following $\Sigma_1(<, +1)$ sentence:

$$\exists x_1 \cdots \exists x_\ell \quad min(x_1) \wedge max(x_\ell) \wedge \left(\bigwedge_{1 \le i \le \ell - 1} x_i + 1 = x_{i+1} \right) \wedge \left(\bigwedge_{1 \le i \le \ell} a_i(x_i) \right).$$

Similarly, A^*w is defined by the following $\Sigma_1(<, +1)$ sentence:

$$\exists x_1 \cdots \exists x_\ell \quad max(x_\ell) \wedge \left(\bigwedge_{1 \le i \le \ell - 1} x_i + 1 = x_{i+1} \right) \wedge \left(\bigwedge_{1 \le i \le \ell} a_i(x_i) \right).$$

Case 2. We now consider languages of the form $\tau_{\mathbf{P}}^{-1}(L_P)$. By hypothesis $L_P \subseteq (\mathbf{P} \times A)^*$ is defined by some $\Sigma_n(<)$ sentence Ψ. We exhibit a $\Sigma_n(<, +1)$ sentence defining $\tau_{\mathbf{P}}^{-1}(L_P)$. For this, we first make sure that all atomic formulas of the form $(Q, a)(x)$ occurring in Ψ are under no negation. This can be assumed since if $(Q, a)(x)$ is such an atomic formula, we have $(Q, a)(x) = \bigvee_{(Q', a') \neq (Q, a)} (Q', a')(x)$.

There are now two sub-cases, depending on whether n is odd or even. If n is odd, then the innermost block of quantifiers is an existential one. Therefore, replacing an atomic sub-formula $(Q, a)(x)$ which is not negated within a $\Sigma_n(<, +1)$ sentence (and in particular within a $\Sigma_n(<)$ sentence such as Ψ) by some $\Sigma_1(<, +1)$ formula yields a $\Sigma_n(<, +1)$ sentence again. We now use the following simple result.

FACT 6.4. *Given any $(Q, a) \in \mathbf{P} \times A$, there exists a $\Sigma_1(<, +1)$ formula $\zeta_{(Q, a)}(x)$ over A with one free variable such that for any $w \in A^*$ and any position x in w, we have $w \models \zeta_{(Q, a)}(x)$ if and only if x has label (Q, a) in $\tau_{\mathbf{P}}(w)$.*

The proof of Fact 6.4 is left to the reader (it is similar to that of Case 1 above). Recall that we have a $\Sigma_n(<)$ sentence Ψ defining L_P. Consider the $\Sigma_n(<, +1)$ sentence φ obtained from Ψ by replacing any atomic formula of the form $(Q, a)(x)$ (for $(Q, a) \in \mathbf{P} \times A$) by the formula $\zeta_{(Q, a)}(x)$ given by Fact 6.4. Then, φ is $\Sigma_n(<, +1)$ defining $\tau_{\mathbf{P}}^{-1}(L_P)$. This concludes the proof for this sub-case.

We now assume that n is even. In that case, replacing an atomic sub-formula under no negation within a $\Sigma_n(<, +1)$ sentence (and in particular within a $\Sigma_n(<)$ sentence such as Ψ) by some $\Pi_1(<, +1)$ formula yields a $\Sigma_n(<, +1)$ sentence. We shall need the following simple result.

FACT 6.5. *Given any $(Q, a) \in \mathbf{P} \times A$, there exists a $\Pi_1(<, +1)$ formula $\xi_{(Q,a)}(x)$ over A with one free variable such that for any $w \in A^*$ and any position x in w, we have $w \models \xi_{(Q,a)}(x)$ if and only if x has label (Q, a) in $\tau_{\mathbf{P}}(w)$.*

PROOF. By Lemma 2.11, $Q \in \mathbf{P}$ is a finite union of languages $\{u\}$ or A^*u where $u \in A^*$. Since $\Pi_1(<, +1)$ is closed under union, it suffices to consider the cases when Q is one of these two kinds of language. If $u = \varepsilon$, then we let $\xi_{(\{\varepsilon\}, a)}(x) = min(x) \wedge a(x)$ and $\xi_{(A^*, a)}(x) = a(x)$.

Otherwise, there exist $a_1, \ldots, a_\ell \in A$ such that $w = a_1 \cdots a_\ell$. Observe that for any $m \in \mathbb{N}$ we have a $\Pi_1(<, +1)$ sentence $\chi_m(x)$ which holds when $x \geq m + 1$. Indeed, we may define this formula by induction on m. When $m = 0$, then $\chi_0(x) = \top$. Otherwise, $\chi_m(x) = \neg min(x) \wedge \forall y \, (y + 1 = x \Rightarrow \chi_{m-1}(y))$. We may now define, $\xi_{(Q,a)}(x)$. If $Q = \{w\}$, we define $\xi_{(Q,a)}(x)$ as the following formula:

$$a(x) \wedge \chi_\ell(x) \wedge \forall x_1 \cdots \forall x_\ell \left(\bigwedge_{i \leq \ell - 1} x_i + 1 = x_{i+1} \wedge x_\ell + 1 = x \right) \Rightarrow \left(\bigwedge_{i \leq \ell} a_i(x_i) \wedge min(x_1) \right).$$

Finally, if $Q = A^*w$, we define $\xi_{(Q,a)}(x)$ as the following formula,

$$a(x) \wedge \chi_\ell(x) \wedge \forall x_1 \cdots \forall x_\ell \left(\bigwedge_{i \leq \ell - 1} x_i + 1 = x_{i+1} \wedge x_\ell + 1 = x \right) \Rightarrow \left(\bigwedge_{i \leq \ell} a_i(x_i) \right).$$

This concludes the proof of Fact 6.5. $\qquad\square$

Recall now that we have a $\Sigma_n(<)$ sentence Ψ defining L_P. Consider the $\Sigma_n(<, +1)$ sentence φ obtained from Ψ by replacing any atomic formula of the form $(Q, a)(x)$ with $(Q, a) \in \mathbf{P} \times A$ (*i.e.*, any label test) by the formula $\xi_{(Q,a)}(x)$ given by Fact 6.4. One may verify that φ is $\Sigma_n(<, +1)$ and defines $\tau_{\mathbf{P}}^{-1}(L_P)$ which concludes the proof for this sub-case.

REMARK 6.6. *When n is even, an alternative proof is to first ensure that all atomic sub-formulas of the form $(Q, a)(x)$ (i.e., label tests) are under exactly one negation in Ψ, and to apply Fact 6.4 again.*

6.2 Ehrenfeucht-Fraïssé games

Before turning to the converse direction in Proposition 6.2, let us recall the definition of the Ehrenfeucht-Fraïssé games associated to the levels Σ_n. It is parameterized by an arbitrary signature σ (which we shall instantiate later with the signatures of $\Sigma_n(<)$ and $\Sigma_n(<, +1)$).

Quantifier rank and canonical preorders. As for two variable first-order logic, the link with Ehrenfeucht-Fraïssé games is based on the notion of quantifier rank. Recall that the rank of a first-order sentence is the longest sequence of nested quantifiers in φ.

Using the quantifier rank, we associate a preorder relation to any level $\Sigma_n(\sigma)$ in the quantifier alternation hierarchy. Given two words $w, w' \in A^*$ and $k \in \mathbb{N}$, we write $w \preccurlyeq_{n,k}^\sigma w'$ when

$$\text{For any } \Sigma_n(\sigma) \text{ sentence of rank at most } k: \quad w \models \varphi \Rightarrow w' \models \varphi.$$

The next lemma is folklore and simple to verify. It characterizes with the preorder $\preccurlyeq_{n,k}^\sigma$ the languages that can be defined by a $\Sigma_n(\sigma)$ sentence of rank k. An *upper set* for $\preccurlyeq_{n,k}^\sigma$ is a language $L \subseteq A^*$ which is upward closed under $\preccurlyeq_{n,k}^\sigma$: given any $w, w' \in A^*$, if $w \in L$ and $w \preccurlyeq_{n,k}^\sigma w'$, then $w' \in L$.

LEMMA 6.7 (FOLKLORE). *Consider two natural numbers $n \geq 1$ and $k \geq 0$. For any language $L \subseteq A^*$, the following two properties are equivalent:*

- *L can be defined by a $\Sigma_n(\sigma)$ sentence of rank k.*
- *L is an upper set for $\preccurlyeq_{n,k}^\sigma$.*

We now define the Ehrenfeucht-Fraïssé game for $\Sigma_n(\sigma)$ (called the $\Sigma_n(\sigma)$ game). It yields an alternate (and easier to manipulate) definition of the preorders $\preccurlyeq_{n,k}^{\sigma}$. The board of the game consists of two words w and w' in A^* and there are two players called *Spoiler* and *Duplicator*. We speak of the $\Sigma_n(\sigma)$-game over the pair (w, w'), or over w and w'. Note that unlike in the FO^2 game, the ordering between the two words is relevant: w is the first word and w' is the second. Spoiler's goal is to prove that the words w and w' are different (wrt. $\Sigma_n(<)$ or $\Sigma_n(<, +1)$) while Duplicator must prevent him from doing so. The game is set to last a predefined number k of rounds and when it starts, each player owns k pebbles. Moreover, we have the two following additional parameters that may change as the play progresses:

(1) There is a distinguished word among w, w', called the *active word*. Initially, the active word is the first word, that is, w.

(2) There is a counter c called the *alternation counter*. Initially, c is set to 0. It can only increase, and its maximal allowed value is $n - 1$. It counts the number of times the active word was changed.

A single round is played as follows. Spoiler has to place a pebble on the board (*i.e.*, on a position of either w or w'). However, there are constraints on the word that he may choose. Spoiler can always choose the active word, in which case both c and the active word remain unchanged. On the other hand, Spoiler may choose the word that is not active only when $c < n - 1$. In that case, the active word is switched and c is incremented by 1.

Duplicator must answer by placing one of her own pebbles on some position of the other word. This answer must yield a *correct configuration*. By *configuration after round* ℓ, we mean the set

$$C = \{(x_1, x_1'), \ldots, (x_\ell, x_\ell')\},$$

where the elements (x_i, x_i') are the pairs of positions (x_i in w and x_i' in w') holding corresponding pebbles at rounds $1, \ldots, \ell$ (*i.e.*, Spoiler placed a pebble on x_i in a previous round and Duplicator answered by putting a pebble on x_i', or *vice versa*). Such a configuration is declared *correct* if and only if for any predicate $P \in \sigma$ of arity m, given any $i_1, \ldots, i_m \leq \ell$,

$$P(x_{i_1}, \ldots, x_{i_m}) \text{ holds} \quad \text{if and only if} \quad P(x_{i_1}', \ldots, x_{i_m}') \text{ holds}.$$

Intuitively, a configuration is correct when it is impossible to point out a difference between the sequences of positions $x_1, \ldots, x_\ell$ in w and $x_1', \ldots, x_\ell'$ in w' by using the predicates available in σ.

When the game starts, the configuration is empty. *Duplicator wins* if this initial configuration is correct (while empty, the initial configuration may not be correct when σ contains constants such as "ε") and if she is able to answer all moves by Spoiler with a correct configuration until all k rounds have been played. On the other hand *Spoiler wins* if the initial configuration is not correct or as soon as Duplicator is unable to play. We now state the Ehrenfeucht-Fraïssé theorem for the $\Sigma_n(\sigma)$-game. It characterizes the preorder $\preccurlyeq_{n,k}^{\sigma}$: two words are comparable, *i.e* $w \preccurlyeq_{n,k}^{\sigma} w'$, when Duplicator has a winning strategy for k rounds over (w, w').

THEOREM 6.8 (FOLKLORE). *Let* $n \geq 1$, $k \in \mathbb{N}$ *and* $w, w' \in A^*$. *Then* $w \preccurlyeq_{n,k}^{\sigma} w'$ *if and only if Duplicator has a winning strategy for playing k rounds in the $\Sigma_n(\sigma)$ game over* (w, w').

This concludes the definition of Ehrenfeucht-Fraïssé games. Note that while the above presentation is generic to all signatures, we are only interested in two specific ones. Given $n \geq 1$ and $k \in \mathbb{N}$, we write $\preccurlyeq_{n,k}$ for the preorder associated to the $\Sigma_n(<)$ sentences of rank k and $\preccurlyeq_{n,k}^+$ for the preorder associated to the $\Sigma_n(<, +1)$ sentence of rank k. Finally, we shall need the following simple result about the relations $\preccurlyeq_{n,k}$.

LEMMA 6.9. *Let* $n \geq 1$ *and let* $h, k \in \mathbb{N}$ *be natural numbers. Consider three words* $w, w', u \in A^*$ *such that* $|u| \leq h$ *and* $wu \preccurlyeq_{n,k+h} w'u$. *Then, we have* $w \preccurlyeq_{n,k} w'$.

PROOF. By hypothesis, we know than Duplicator has a winning strategy for playing $k + h$ rounds in the $\Sigma_n(<)$-game over wu and $w'u$. Since $|u| \leq h$, it is simple to verify that as long as there are more than h rounds remaining after the current one, if Spoiler places a pebble in one of the prefixes w or w', then Duplicator's strategy gives an answer in w or w'. Otherwise, Spoiler would be able to win within the h following rounds (it is important here that the signature of $\Sigma_n(<)$ includes the linear order "$<$"). Therefore, it is immediate that Duplicator gets a winning strategy for playing k rounds in the $\Sigma_n(<)$-game over w and w'. This means that $w \preccurlyeq_{n,k} w'$, as desired. □

6.3 From successor to enrichment

We are now ready to prove the remaining direction in Proposition 6.2. For any $n \geq 1$, we show that $\Sigma_n(<, +1) \subseteq \Sigma_n(<) \circ \mathrm{SU}$.

Let us start with some preliminary definitions. Recall that for any $k \in \mathbb{N}$, we denote by $\sim_k$ the canonical equivalence on A^* associated to SU_k. We write $\mathbf{P}_k$ for the finite partition of A^* into $\sim_k$-classes. Recall that $\mathbf{P}_k$ is an SU-partition of A^* by Lemma 2.10. We shall only use the SU-partitions $\mathbf{P}_k$ to build languages in $\Sigma_n(<) \circ \mathrm{SU}$. For the sake of simplifying the notation, given $k \in \mathbb{N}$, we write

- τ_k for the map $\tau_{\mathbf{P}_k} : A^* \longrightarrow (\mathbf{P}_k \times A^*)$.
- δ_k for the map $\delta_{\mathbf{P}_k} : A^* \times A^* \longrightarrow (\mathbf{P}_k \times A^*)$.

We now prove that $\Sigma_n(<, +1) \subseteq \Sigma_n(<) \circ \mathrm{SU}$. Our argument is based on the following proposition.

PROPOSITION 6.10. *Let $k \geq 0$, $n \geq 1$ and $\ell = 2^k$ be three integers. Assume that we have $w, w' \in A^*$ such that $\tau_\ell(w) \preccurlyeq_{n,k+\ell} \tau_\ell(w')$ and $w \sim_\ell w'$. Then, we have $w \preccurlyeq_{n,k}^+ w'$.*

Before we show Proposition 6.10, let us use it to conclude this direction of the proof. Let $L \subseteq A^*$ be a language defined by some $\Sigma_n(<, +1)$ sentence φ. We show that $L \in \Sigma_n(<) \circ \mathrm{SU}$. By Lemma 6.7, L is an upper set for the preorder $\preccurlyeq_{n,k}^+$, where k is the rank of φ.

Let $\ell = 2^k$. For any $P \in \mathbf{P}_\ell$, we let $H_P \subseteq (\mathbf{P}_\ell \times A)^*$ be the following upper set for $\preccurlyeq_{n,k+\ell}$:

$$H_P = \{u \in (\mathbf{P}_\ell \times A)^* \mid \text{there exists } w \in P \cap L \text{ such that } \tau_\ell(w) \preccurlyeq_{n,k+\ell} u\}.$$

Since H_P is an upper set for $\preccurlyeq_{n,k+\ell}$, Lemma 6.7 entails that it can be defined by a $\Sigma_n(<)$ sentence (of rank $k + \ell$). To conclude, we will show that

$$L = \bigcup_{P \in \mathbf{P}_\ell} (P \cap \tau_\ell^{-1}(H_P)). \tag{5}$$

It will then be immediate that $L \in \Sigma_n(<) \circ \mathrm{SU}$, since $\mathbf{P}_\ell$ is an SU-partition of A^*.

It remains to prove (5). We start with the left to right inclusion. Assume that $v \in L$. Since $\mathbf{P}_\ell$ is a partition of A^*, there exists some unique $P \in \mathbf{P}_\ell$ such that $v \in P$. It is then immediate from the definition that $\tau_\ell(v) \in H_P$ since $\tau_\ell(v) \preccurlyeq_{n,k+\ell} \tau_\ell(v)$. Thus, we get $v \in P \cap \tau_\ell^{-1}(H_P)$ which concludes the proof of this inclusion.

We turn to the right to left inclusion. Assume that $v \in P \cap \tau_\ell^{-1}(H_P)$ for some $P \in \mathbf{P}_\ell$. We want to show that $v \in L$. Since $v \in \tau_\ell^{-1}(H_P)$, we obtain by definition of H_P some word $w \in P \cap L$ such that $\tau_\ell(w) \preccurlyeq_{n,k+\ell} \tau_\ell(v)$. Moreover, since v and w both belong to P, we have $w \sim_\ell v$. Thus, since $\ell = 2^k$ by definition, we obtain the relation $w \preccurlyeq_{n,k}^+ v$ from Proposition 6.10. Finally, since $w \in L$ and since L is an upper set for $\preccurlyeq_{n,k}^+$, we get $v \in L$, as desired.

It remains to prove Proposition 6.10, to which we devote the end of the section. Let $k \geq 0$, $n \geq 1$ and $\ell = 2^k$. Consider two words $w, w' \in A^*$ such that $\tau_\ell(w) \preccurlyeq_{n,k+\ell} \tau_\ell(w')$ and $w \sim_\ell w'$. We prove that $w \preccurlyeq_{n,k}^+ w'$. As expected, we use an Ehrenfeucht-Fraïssé argument and describe a winning

strategy for Duplicator in the $\Sigma_n(<, +1)$-game over w and w'. Recall that there are k rounds to play, that the alternation counter c starts at 0 and has to remains bounded by $n - 1$. We use an induction on n and k (in any order) to describe Duplicator's winning strategy.

Assume first that $k = 0$, which means that $\ell = 1$. In that case, there are no rounds to play and it suffices to show that Duplicator wins automatically (*i.e.*, that w and w' satisfy the same constants in the signature of $\Sigma_n(<, +1)$). There is only one constant in the signature of $\Sigma_n(<, +1)$: "ε". It is immediate that $w \models \varepsilon$ if and only if $w' \models \varepsilon$ since we know that $w \sim_1 w'$ by hypothesis. This concludes the case $k = 0$.

We now assume that $k \geq 1$. We need to describe a strategy for Duplicator in order to play k rounds in the $\Sigma_n(<, +1)$-game over w and w'. Consider a move by Spoiler in the first round. We show that Duplicator is able to answer this move and then to win the remaining $k - 1$ rounds. The argument depends on whether Spoiler plays his first move in w or in w'. If Spoiler plays in w', we use induction on n. In that case, the alternation counter is incremented (in particular, this may only happen when $n \geq 2$). One may verify from the definition that the game now corresponds to a $\Sigma_{n-1}(<, +1)$-game over (w', w). Hence, it suffices to show that Duplicator has a winning strategy for playing k rounds in this simpler game. This is immediate from induction on n. Indeed, we know that $\tau_\ell(w) \preccurlyeq_{n,k+\ell} \tau_\ell(w')$ and $w \sim_\ell w'$ by hypothesis. One may verify that this implies $\tau_\ell(w') \preccurlyeq_{n-1,k+\ell} \tau_\ell(w)$ and $w' \sim_\ell w$. Hence, we obtain from induction on n that $w' \preccurlyeq_{n-1,k}^{+1} w$ which yields the desired strategy for Duplicator.

It remains to handle the case when Spoiler plays his first move on some position x of the word w. This requires more work. We may decompose w according to the position x: $w = uav$ where the highlighted letter a is at position x. We use the following lemma to describe an answer for Duplicator.

LEMMA 6.11. *The word w' has a decomposition $w' = u'av'$ such that $u \preccurlyeq_{n,k-1}^+ u'$ and $v \preccurlyeq_{n,k-1}^+ v'$.*

Lemma 6.11 provides Duplicator's answer to Spoiler's first move: consider the decomposition $w' = u'av'$ given by the lemma and let x' be the position of w' corresponding to the highlighted letter a in this decomposition. We choose x' as Duplicator's answer. One may verify that this answer is correct (*i.e.*, x and x' satisfy the same predicates in the signature of $\Sigma_n(<, +1)$).

REMARK 6.12. *For showing that $min(x)$ holds if and only if $min(x')$, one needs to use the fact that $u \preccurlyeq_{n,k-1}^+ u'$ (which means that $u = \varepsilon$ if and only if $u' = \varepsilon$). Symmetrically, the fact that $max(x)$ holds if and only if $max(x')$ is based on $v \preccurlyeq_{n,k-1}^+ v'$.*

It remains to show that Duplicator has a winning strategy for playing $k - 1$ more rounds in the $\Sigma_n(<, +1)$-game over $w = uav$ and $w' = u'av'$ from the configuration $C = \{(x, x')\}$. This can be verified using our hypothesis that $u \preccurlyeq_{n,k-1}^+ u'$ and $v \preccurlyeq_{n,k-1}^+ v'$. Indeed, by induction, we get strategies for playing $k - 1$ rounds over u and u', and v and v' respectively. These strategies are easily combined into a single one for playing $k - 1$ rounds over $w = uav$ and $w' = u'av'$.

We finish with the proof of Lemma 6.11. Recall that we have $w = uav$ where the highlighted letter "a" is at position x. We consider two cases depending on the length of v (*i.e.*, on whether "x" is close to the "right border" of w). We let $h = 2^{k-1}$. Note that by definition, we have $\ell = 2h$.

First case. Assume first $|v| < h$. In that case, av is a suffix of length at most h of w. Since we know that $w \sim_\ell w'$ and $\ell = 2h$ by hypothesis, it follows that av is a suffix of w' as well. In other words, we obtain that w' admits a decomposition $w' = u'av$ for some $u' \in A^*$. It is immediate that $v \preccurlyeq_{n,k-1}^+ v$. It remains to show that $u \preccurlyeq_{n,k-1}^+ u'$.

We prove that $\tau_h(u) \preccurlyeq_{n,k-1+h} \tau_h(u')$ and $u \sim_h u'$. It will then be immediate by induction on k in Proposition 6.10 that we have $u \preccurlyeq^+_{n,k-1} u'$, as desired. We start with the equivalence $u \sim_h u'$. Recall that by hypothesis we have,

$$uav = w \sim_\ell w' = u'av.$$

Thus, uav and $u'av$ have the same suffixes of length at most $\ell = 2h$. Moreover, since $|av| \leq h$, it is immediate that u and u' have the same suffixes of length at most h, which means that $u \sim_h u'$.

It remains to show that $\tau_h(u) \preccurlyeq_{n,k-1+h} \tau_h(u')$. By hypothesis, we have, $\tau_\ell(w) \preccurlyeq_{n,k+\ell} \tau_\ell(w')$. Since $\ell \geq h$, this implies $\tau_h(w) \preccurlyeq_{n,k+\ell} \tau_h(w')$. Moreover, we have $w = uav$ and $w' = u'av$. Therefore, by Lemma 4.4:

$$\tau_h(u) \cdot \delta_h(u, av) \preccurlyeq_{n,k+\ell} \tau_h(u') \cdot \delta_h(u', av)$$

We just proved that $u \sim_h u'$, which, together with the definition of δ_h, implies $\delta_h(u, av) = \delta_h(u', av)$. Altogether, this means that there exists $z \in (\mathbf{P}_h \times A)^*$ such that $|z| \leq h$ and,

$$\tau_h(u) \cdot z \preccurlyeq_{n,k+\ell} \tau_h(u') \cdot z$$

It now follows from Lemma 6.9 that $\tau_h(u) \preccurlyeq_{n,k+\ell-h} \tau_h(u')$, and since $\ell = 2h$, we get in particular $\tau_h(u) \preccurlyeq_{n,k-1+h} \tau_h(u')$, as desired.

Second case. We now assume that $|v| \geq h$. Since the highlighted "a" in $w = uav$ is at position x, it follows by hypothesis on v that $y = x + h$ is also a position of w. Hence, we may further decompose w according to y: $w = uav_1bv_2$, where the highlighted b is at position y. In other words $v = v_1bv_2$. Observe that by definition $v_1 = w[x + 1, y - 1]$ which means that $|av_1| = h$. We use the following fact, whose proof relies on our hypothesis that $\tau_\ell(w) \preccurlyeq_{n,k+\ell} \tau_\ell(w')$.

FACT 6.13. *There exists $u', v'_2 \in A^*$ such that $w' = u'av_1bv'_2$ and the following properties are satisfied:*

- $uav_1 \sim_\ell u'av_1$.
- $\tau_\ell(uav_1) \preccurlyeq_{n,k+\ell-1} \tau_\ell(u'av_1)$.
- $\delta_\ell(uav_1b, v_2) \preccurlyeq_{n,k+\ell-1} \delta_\ell(u'av_1b, v'_2)$.

PROOF. By definition, w and $\tau_\ell(w)$ share the same set of positions. Thus, we may view x and y as positions in $\tau_\ell(w)$. In particular, we get from Lemma 4.4 that:

$$\tau_\ell(w) = \tau_\ell(uav_1) \cdot ([uav_1]_{\mathbf{P}_\ell}, b) \cdot \delta_\ell(uav_1b, v_2).$$

Since $\tau_\ell(w) \preccurlyeq_{n,k+\ell} \tau_\ell(w')$, Duplicator has a winning strategy for playing $(k + \ell)$ rounds in the $\Sigma_n(<)$-game over $\tau_\ell(w)$ and $\tau_\ell(w')$. She may simulate a move by Spoiler in this game by placing a pebble on the position y in $\tau_\ell(w)$. Her strategy then yields an answer y' in $\tau_\ell(w)$. Recall that we may view y' as a position of w'. We decompose w' as $w' = z'cv'_2$ where the highlighted letter $c \in A$ is at position y'. It then follows from Lemma 4.4 that,

$$\tau_\ell(w') = \tau_\ell(z') \cdot ([z']_{\mathbf{P}_\ell}, c) \cdot \delta_\ell(z'c, v'_2).$$

By definition of y', we know that y and y' have the same label in $\tau_\ell(w)$ and $\tau_\ell(w')$. Thus, it is immediate that $b = c$ and $[uav_1]_{\mathbf{P}_\ell} = [z']_{\mathbf{P}_\ell}$. Note that by definition, the latter property means that $uav_1 \sim_\ell z'$. In particular, since $|av_1| = h \leq \ell$, it follows that we have $z' = u'av_1$ for some $u' \in A^*$. Altogether, we have found a decomposition $w' = u'av_1bv'_2$ with $uav_1 \sim_\ell u'av_1$.

Moreover, we know that Duplicator has a strategy for playing $k + \ell - 1$ more rounds in the $\Sigma_n(<)$-game over $\tau_\ell(w)$ and $\tau_\ell(w')$ from the configuration $\{(y, y')\}$. It follows that $\tau_\ell(uav_1) \preccurlyeq_{n,k+\ell-1} \tau_\ell(u'av_1)$ and $\delta_\ell(uav_1b, v_2) \preccurlyeq_{n,k+\ell-1} \delta_\ell(u'av_1b, v'_2)$ which concludes the proof. □

We may now come back to the proof of Case 2 and describe our decomposition of w'. We let $w' = u'av_1bv_2'$ be the decomposition given by Fact 6.13. Finally, we define $v' = v_1bv_2'$. We now have our decomposition $w' = u'av'$. It remains to show that $u \preccurlyeq_{n,k-1}^+ u'$ and $v \preccurlyeq_{n,k-1}^+ v'$.

Let us start with $u \preccurlyeq_{n,k-1}^+ u'$. We prove that $\tau_h(u) \preccurlyeq_{n,k-1+h} \tau_h(u')$ and $u \sim_h u'$. It will then be immediate from induction on k in Proposition 6.10 that $u \preccurlyeq_{n,k-1}^+ u'$ as desired. For the equivalence $u \sim_h u'$, we know from the first item in Fact 6.13 that,

$$uav_1 \sim_\ell u'av_1$$

Thus, since $|av_1| = h$ and $\ell \geq 2h$, it is immediate that $u \sim_h u'$. We turn to $\tau_h(u) \preccurlyeq_{n,k-1+h} \tau_h(u')$. By the second item in Fact 6.13 we have:

$$\tau_\ell(uav_1) \preccurlyeq_{n,k+\ell-1} \tau_\ell(u'av_1).$$

Since $\ell \geq h$, one may verify that this implies $\tau_h(uav_1) \preccurlyeq_{n,k+\ell-1} \tau_h(uav_1')$. Using Lemma 4.4, we obtain

$$\tau_h(u) \cdot \delta_h(u, av_1) \preccurlyeq_{n,k+\ell-1} \tau_h(u') \cdot \delta_h(u', av_1).$$

Moreover, since $u \sim_h u'$, the definition of δ_h entails that $\delta_h(u, av_1) = \delta_h(u', av_1)$. Let $z = \delta_h(u, av_1)$. By definition, we have $|z| \leq h$ and,

$$\tau_h(u) \cdot z \preccurlyeq_{n,k-1+\ell} \tau_h(u') \cdot z.$$

Since $\ell \geq 2h$, we have $k - 1 + \ell \geq k - 1 + h + h$ and it now follows from Lemma 6.9 that $\tau_h(u) \preccurlyeq_{n,k-1+h} \tau_h(u')$, as desired.

We finish with the inequality $v \preccurlyeq_{n,k-1}^+ v'$. We reuse the same approach, by showing that $\tau_h(v) \preccurlyeq_{n,k-1+h} \tau_h(v')$ and $v \sim_h v'$. The result will then follow by the induction on k in the proof of Proposition 6.10. Recall that $v = v_1bv_2$ and $v' = v_1bv_2'$.

For proving that $v \sim_h v'$, recall that by hypothesis we have $w \sim_\ell w'$ with $\ell = 2h$. Thus, we have $w \sim_h w'$. Moreover, $|v_1b| = h$ by hypothesis. Thus, $v = v_1bv_2$ and $v' = v_1bv_2'$ are suffixes of w and w' of length larger than h. Altogether, it follows that $v \sim_h v'$.

It remains to show that, $\tau_h(v) \preccurlyeq_{n,k-1+h} \tau_h(v')$. Since $v = v_1bv_2$ and $v' = v_1bv_2'$, we get from Lemma 4.4 that:

$$\tau_h(v_1bv_2) = \tau_h(v_1) \cdot \delta_h(v_1b, v_2) \quad \text{and} \quad \tau_h(v_1bv_2') = \tau_h(v_1) \cdot \delta_h(v_1b, v_2').$$

It is straightforward to verify that $\preccurlyeq_{n,k-1+h}$ is compatible with concatenation. Since clearly, $\tau_h(v_1) \preccurlyeq_{n,k-1+h} \tau_h(v_1)$, it suffices to show that we have $\delta_h(v_1b, v_2) \preccurlyeq_{n,k-1+h} \delta_h(v_1b, v_2')$. By the third item in Fact 6.13, we have

$$\delta_\ell(uav_1b, v_2) \preccurlyeq_{n,k-1+\ell} \delta_\ell(u'av_1b, v_2').$$

Since $\ell \geq h$, one may verify that this implies $\delta_h(uav_1b, v_2) \preccurlyeq_{n,k+\ell-1} \delta_h(u'av_1b, v_2')$. Moreover, since $|v_1b| = h$, the definition of δ_h gives us that $\delta_h(uav_1b, v_2) = \delta_h(v_1b, v_2)$ and $\delta_h(u'av_1b, v_2') = \delta_h(v_1b, v_2')$. Therefore, we obtain:

$$\delta_h(v_1b, v_2) \preccurlyeq_{n,k-1+\ell} \delta_h(v_1b, v_2').$$

In particular, this implies $\delta_h(v_1b, v_2) \preccurlyeq_{n,k-1+h} \delta_h(v_1b, v_2')$, which concludes the proof.

7 THE REDUCTION FOR ω-WORDS

In this section, we generalize our reduction to the setting of ω-languages. We follow the same outline as the one we used for languages of finite words in Section 4. First, we adapt the definition of enrichment to classes of ω-languages. In this setting, enrichment combines objects of different nature: given a class of ω-languages C and a class of languages $\mathcal{D}$ (such as SU), we define the $\mathcal{D}$-enrichment of C (still denoted by $C \circ \mathcal{D}$). Then, we generalize the reduction theorem (Theorem 4.12): given any lattice of ω-languages C closed under inverse image, $(C \circ \mathrm{SU})$-covering reduces to C-covering.

REMARK 7.1. *Our statements and proofs in this section are very similar to the ones we presented for words in Section 4. In fact, aside from one specific technical result, our main theorem for ω-words and its proof are both straightforward generalizations of Theorem 4.12. For this reason, we shall often leave the proofs of technical sub-results to the reader and refer to the corresponding statement in Section 4.*

7.1 Enrichment for classes of ω-languages

We generalize enrichment to classes of ω-languages. Let us first adapt **P**-taggings.

P-taggings. Let A be an alphabet and **P** a finite partition of A^*. We define a canonical map $\tau_\mathbf{P} : A^\omega \to (\mathbf{P} \times A)^\omega$. Let $w \in A^\omega$ be an ω-word: $w = a_1 a_2 a_3 \cdots$ with $a_i \in A$ for all i. We let $\tau_\mathbf{P}(w)$ be the ω-word $\tau_\mathbf{P}(w) = b_1 b_2 b_3 \cdots$ where,

$$b_1 = ([\varepsilon]_\mathbf{P}, a_1) \quad \text{and} \quad b_i = ([a_1 \cdots a_{i-1}]_\mathbf{P}, a_i) \quad \text{for } i \geq 2$$

Enrichment. Consider a class of ω-languages C and a class of languages $\mathcal{D}$ (do note that $\mathcal{D}$ is a class of languages and not of ω-languages). The $\mathcal{D}$-enrichment of C, denoted by $C \circ \mathcal{D}$ is now defined as the following class of ω-languages. For any alphabet A, $(C \circ \mathcal{D})(A)$ contains all ω-languages of the following form:

$$\tau_\mathbf{P}^{-1}(L) \quad \text{where } \mathbf{P} \text{ is a } \mathcal{D}\text{-partition of } A^* \text{ and } L \in C(\mathbf{P} \times A).$$

REMARK 7.2. *The definition is actually simpler in this setting. Indeed, since we are dealing with ω-languages, it makes no sense to consider intersections with elements of* **P***, which are word languages.*

As before, we are mainly interested in SU-enrichment since our theorem applies to this special case. As for finite words, SU-enrichment for classes of ω-languages captures the intuitive connection between strong and weak logical fragments. One may show that over ω-words as well, $\mathrm{FO}^2(<, +1)$ is the SU-enrichment of $\mathrm{FO}^2(<)$ and for any $n \geq 1$ $\Sigma_n(<, +1)$ and $\mathcal{B}\Sigma_n(<, +1)$ are respectively the SU-enrichments of $\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<)$. Since the proofs are essentially identical[3] to those we presented in Sections 5 and 6 for finite words, they are left to the reader.

We now turn to the variant for ω-words of our main theorem: given any lattice C of ω-languages which is closed under right quotient and inverse image, $(C \circ \mathrm{SU})$-covering reduces to C-covering. Both the reduction and its proofs are adapted from what we did for finite words in Section 4. We start by generalizing well-formed words.

7.2 Languages of well-formed ω-words

Similarly, to what happened for finite words in Section 4, using our reduction for ω-words requires working with the algebraic definition of regular ω-languages, which is based on ω-semigroups. We first briefly recall the definition of ω-semigroups and refer the reader to the book of Perrin and Pin [20] for more details.

---

[3]In fact, the proofs are even simpler in this setting. Since ω-words have no "right border", there are less cases to treat.

ω-**semigroups.** An ω-*semigroup* is a pair (S_+, S_ω) where S_+ is a semigroup and S_ω is a set. Moreover, (S_+, S_ω) is equipped with two additional products: a *mixed product* $S_+ \times S_\omega \to S_\omega$ that maps $s \in S_+$ and $t \in S_\omega$ to an element denoted $st \in S_\omega$, and an *infinite product* $(S_+)^\omega \to S_\omega$ that maps an infinite sequence $s_1, s_2, \dots \in (S_+)^\omega$ to an element of S_ω denoted by $s_1 s_2 \cdots$. We require these products as well as the semigroup product of S_+ to satisfy all possible forms of associativity (see [20] for details). Finally, we denote by s^ω the element $sss \cdots$. Clearly, (A^+, A^ω) is an ω-semigroup for any alphabet A. The notion of morphism is adapted to ω-semigroups in the natural way.

An ω-semigroup is said to be *finite* if both S_+ and S_ω are finite. Note that even if an ω-semigroup is finite, it is not obvious that a finite representation of the infinite product exists. However, it was proven by Wilke [46] that the infinite product is fully determined by the mapping $s \mapsto s^\omega$, yielding a finite representation for finite ω-semigroups.

An ω-language $L \subseteq A^\omega$ is said to be *recognized* by an ω-semigroup (S_+, S_ω) if there exist $F \subseteq S_\omega$ and a morphism $\alpha : (A^+, A^\omega) \to (S_+, S_\omega)$ such that $L = \alpha^{-1}(F)$. It is well known that an ω-language is regular if and only if it is recognized by a *finite* ω-semigroup.

Well-formed ω-words. We may now adapt the notion of well-formed words to ω-words. To any morphism $\alpha : (A^+, A^\omega) \to (S_+, S_\omega)$ into a finite ω-semigroup, we associate a new alphabet $\mathbb{A}_\alpha$ of *well-formed ω-words*. Then, given any ω-language $L \subseteq A^\omega$ recognized by α, we associate a new ω-language $\mathrm{WF}_\alpha(L) \subseteq \mathbb{A}_\alpha^\omega$.

We denote by S the semigroup $S = \alpha(A^+) \subseteq S_+$. Moreover, we write $E(S)$ for the set of idempotent elements in S. Let "$\square$" be some symbol which does not belong to S. The *alphabet of well-formed ω-words associated to α*, denoted by $\mathbb{A}_\alpha$, is defined as follows:

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

REMARK 7.3. *This definition is simpler than the one for words. We do not need letters of the form $(e, s, \square)$ as ω-words do not have a "right border".*

The definition of well-formed ω-words is the natural one. We say that an ω-word $w \in \mathbb{A}_\alpha^\omega$ is *well-formed* when it can be written as follows:

$$w = (\square, s_0, f_0) \cdot (e_1, s_1, f_1) \cdot (e_2, s_2, f_2) \cdots$$

with $f_i = e_{i+1} \in E(S)$ for all $i \in \mathbb{N}$. It is immediate by definition that the language of all well-formed ω-words in $\mathbb{A}_\alpha^\omega$ is regular.

FACT 7.4. *The language of all well-formed ω-words in $\mathbb{A}_\alpha^\omega$ is regular.*

We now associate a new ω-language over $\mathbb{A}_\alpha$ to each ω-language L recognized by α: the *language of well-formed ω-words* associated to L. As the name suggests, it is made exclusively of well-formed ω-words.

We define a canonical morphism EVAL : $(\mathbb{A}_\alpha^+, \mathbb{A}_\alpha^\omega) \to (S_+, S_\omega)$ by giving the image of the two kinds of letters in $\mathbb{A}_\alpha$. Let $s \in S$ and $e, f \in E(S)$, we define,

$$\mathrm{EVAL}((e, s, f)) = esf \qquad \mathrm{EVAL}((\square, s, f)) = sf.$$

Consider an ω-language L recognized by α. We define $\mathrm{WF}_\alpha(L) \subseteq \mathbb{A}_\alpha^\omega$ as follows:

$$\mathrm{WF}_\alpha(L) = \{w \in \mathbb{A}_\alpha^\omega \mid w \text{ is well-formed and } \mathrm{EVAL}(w) \in \alpha(L)\}.$$

Clearly, $\mathrm{WF}_\alpha(L)$ is the intersection of the language of all well-formed ω-words with an ω-language recognized by EVAL. Thus, it is regular.

FACT 7.5. *For any ω-language L recognized by α, $\mathrm{WF}_\alpha(L)$ is regular.*

Finally, we lift the definition to multisets **L** made of ω-languages recognized by α and write $\mathrm{WF}_\alpha(\mathbf{L})$ for the multiset $\mathrm{WF}_\alpha(\mathbf{L}) = \{\mathrm{WF}_\alpha(L) \mid L \in \mathbf{L}\}$.

7.3 The reduction theorem for ω-words

We may now adapt Theorem 4.12 to ω-words. We state an effective reduction from $(C \circ \mathrm{SU})$-covering to C-covering which holds for any lattice of ω-languages C closed under inverse image (note that unlike in the setting of finite words, we do not require C to be nontrivial here).

THEOREM 7.6. *Let* $\alpha : (A^+, A^\omega) \to (S_+, S_\omega)$ *be an* ω-*semigroup morphism and let* C *be a lattice of* ω-*languages closed under inverse image. Moreover, let* L *be an* ω-*language and* $\mathbf{L}$ *be a multiset of* ω-*languages, all recognized by* α. *Then, the following properties are equivalent:*

(1) $(L, \mathbf{L})$ *is* $(C \circ \mathrm{SU})$-*coverable.*
(2) $(L, \mathbf{L})$ *is* $(C \circ \mathrm{SU}_n)$-*coverable, where* $n = 2^{3|S_+|+1}$.
(3) $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$ *is* C-*coverable.*

As announced, the statement of Theorem 7.6 is a natural analogue of Theorem 4.12. Let us point out that in this case as well, we shall present a constructive proof.

REMARK 7.7. *The constant used in the second item of Theorem 7.6 is much larger than the corresponding one in Theorem 4.12. This is explained by technical difficulties that are specific to* ω-*words and arise when proving the implication* (3) $\Rightarrow$ (2) *of the theorem.*

Finally, note that as before, we may adapt Theorem 7.6 to accommodate the simpler separation problem.

COROLLARY 7.8. *Let* $\alpha : (A^+, A^\omega) \to (S_+, S_\omega)$ *be an* ω-*semigroup morphism and let* C *be a lattice of* ω-*languages closed under inverse image. Moreover, let* L_1, L_2 *be two* ω-*languages recognized by* α. *Then, the following properties are equivalent:*

(1) L_1 *is* $(C \circ \mathrm{SU})$-*separable from* L_2.
(2) L_1 *is* $(C \circ \mathrm{SU}_n)$-*separable from* L_2 *where* $n = 2^{3|S_+|+1}$.
(3) $\mathrm{WF}_\alpha(L_1)$ *is* C-*separable from* $\mathrm{WF}_\alpha(L_2)$.

While we shall not detail the applications of Theorem 7.6 as much as we did for Theorem 4.12, let us briefly outline them. As we explained above, one may show that over ω-words, $\mathrm{FO}^2(<, +1)$ is the SU-enrichment of $\mathrm{FO}^2(<)$ and for any $n \geq 1$ $\Sigma_n(<, +1)$ and $\mathcal{B}\Sigma_n(<, +1)$ are respectively the SU-enrichments of $\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<)$. It was shown in [21] that over ω-words, separation is decidable for the levels $\Sigma_2(<)$ and $\Sigma_3(<)$ of the quantifier alternation hierarchy. Thus, we obtain from Theorem 7.6 that separation is decidable for $\Sigma_2(<, +1)$ and $\Sigma_3(<, +1)$ over ω-words.

REMARK 7.9. *We do not speak about covering since there are no published results for covering in the setting of* ω-*words. We also do not mention separation* $\mathrm{FO}^2(<, +1)$, $\Sigma_1(<, +1)$ *and* $\mathcal{B}\Sigma_1(<, +1)$ *over* ω-*words for the same reason. However, let us point out that our problem here is just the lack of bibliography. It is actually possible to generalize the existing results and show that covering and separation are both decidable for* $\mathrm{FO}^2(<)$, $\Sigma_1(<)$, $\mathcal{B}\Sigma_1(<)$, $\Sigma_2(<)$, $\mathcal{B}\Sigma_2(<)$ *and* $\Sigma_3(<)$ *overs* ω-*words. Thus, Theorem 7.6 can be used to obtain the same results over* ω-*words as the ones obtained from Theorem 4.12 over words.*

The remainder of this section is devoted to proving Theorem 7.6. We fix an arbitrary morphism $\alpha : (A^+, A^\omega) \to (S_+, S_\omega)$ and we let S be the semigroup $S = \alpha(A^+)$. Recall that the associated alphabet of well-formed words is defined as follows:

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

Let C be a lattice of ω-languages closed under inverse image. Our objective is to show that when L and $\mathbf{L}$ are respectively a ω-language and a multiset of ω-languages, all recognized by α, the following properties are equivalent:

(1) $(L, \mathbf{L})$ is $(C \circ \mathrm{SU})$-coverable.

(2) $(L, \mathbf{L})$ is $(C \circ \mathrm{SU}_n)$-coverable where $n = 2^{3|S_+|+1}$.

(3) $(\mathrm{WF}_\alpha(L), \mathrm{WF}_\alpha(\mathbf{L}))$ is C-coverable.

We prove that $(1) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1)$. As before, observe that the direction $(2) \Rightarrow (1)$ is trivial since it is clear that $C \circ \mathrm{SU}_n \subseteq C \circ \mathrm{SU}$. Thus, we may concentrate on $(1) \Rightarrow (3)$ and $(3) \Rightarrow (2)$.

The argument for the direction $(1) \Rightarrow (3)$ is basically identical to the one we used when proving the corresponding implication in Theorem 4.12. For this reason, we shall only briefly sketch it. On the other hand, we provide more details for the proof of the implication $(3) \Rightarrow (2)$, which slightly departs from what we did for words.

7.4 From $(C \circ \mathrm{SU})$-covering to C-covering

We start with the direction $(1) \Rightarrow (3)$. As we explained above, the argument is essentially the same as the one we presented for the corresponding direction in Theorem 4.12. In fact, it is even made simpler by the fact that the definition of $C \circ \mathrm{SU}$ is less involved for classes of ω-languages.

The argument is based on the following proposition which is adapted from the one we used to handle the corresponding direction for words: Proposition 4.15.

PROPOSITION 7.10. *For any integer $k \geq 1$, there exists a map $\gamma : \mathbb{A}_\alpha^\omega \to A^\omega$ satisfying the two following properties:*

(1) *For any ω-language $L \subseteq A^\omega$ recognized by α and any well-formed ω-word $w \in \mathbb{A}_\alpha^\omega$, we have $w \in \mathrm{WF}_\alpha(L)$ if and only if $\gamma(w) \in L$.*

(2) *For any ω-language $K \in (C \circ \mathrm{SU}_k)(A)$, there exists $H_K \in C(\mathbb{A}_\alpha)$ such that for any well-formed ω-word $w \in \mathbb{A}_\alpha^\omega$, $w \in H_K$ if and only if $\gamma(w) \in K$.*

One may show the direction $(1) \Rightarrow (3)$ in Theorem 7.6 from this proposition using an argument which is similar to the one used for proving the same direction in Theorem 4.12 from Proposition 4.15. We leave this argument to the reader and prove Proposition 7.10. Let us fix $k \geq 1$ for the proof.

Proof of Proposition 7.10: definition of γ. We start by defining the map $\gamma : \mathbb{A}_\alpha^\omega \to A^\omega$ and then show that it satisfies the desired properties. We actually define a morphism $\gamma : \mathbb{A}_\alpha^+ \to A^+$ which we lift as a map $\gamma : \mathbb{A}_\alpha^\omega \to A^\omega$. Hence, it suffices to describe the image of any letter in $\mathbb{A}_\alpha$.

For any element $s \in S$ (recall that $S = \alpha(A^+)$), we associate an arbitrarily chosen **nonempty** word $\lceil s \rceil \in A^+$ such that $\alpha(\lceil s \rceil) = s$ (note that such a word exists by definition of S). We are now ready to define our morphism $\gamma : \mathbb{A}_\alpha^+ \to A^+$. Recall that there are two kinds of letters in $\mathbb{A}_\alpha$. Given $s \in S$ and $e, f \in E(S)$, we define,

$$\begin{aligned} \gamma((e, s, f)) &= \lceil e \rceil^k \lceil s \rceil \lceil f \rceil^k \\ \gamma((\square, s, f)) &= \lceil s \rceil \lceil f \rceil^k \end{aligned}$$

Given $w = b_1 b_2 \cdots \in \mathbb{A}_\alpha^\omega$, we now define $\gamma(w) = \gamma(b_1)\gamma(b_2) \cdots \in A^\omega$. It remains to prove that γ satisfies the two properties stated in Proposition 4.15.

Proof of Proposition 7.10: first item. Consider an ω-language $L \subseteq A^\omega$ which is recognized by α. We have to show that for any well-formed ω-word $w \in \mathbb{A}_\alpha^\omega$, $w \in \mathrm{WF}_\alpha(L)$ if and only if $\gamma(w) \in L$.

Since w is well-formed, we have $w \in \mathrm{WF}_\alpha(L)$ if and only if $\mathrm{EVAL}(w) \in \alpha(L)$. Moreover, since α recognizes L, we have $\gamma(w) \in L$ if and only if $\alpha(\gamma(w)) \in \alpha(L)$. Hence, it suffices to prove that $\mathrm{EVAL}(w) = \alpha(\gamma(w))$. By definition,

$$\begin{aligned} w &= (\square, s_0, e_1) \cdot (e_1, s_1, e_2) \cdot (e_2, s_2, e_3) \cdots \\ \gamma(w) &= \lceil s_0 \rceil \lceil e_1 \rceil^{2k} \lceil s_1 \rceil \lceil e_2 \rceil^{2k} \lceil s_2 \rceil \cdot \lceil e_3 \rceil^{2k} \cdots \end{aligned}$$

Hence, we have,

$$
\begin{array}{rcl}
\textsc{eval}(w) & = & s_0 e_1 e_1 s_1 e_2 s_2 e_3 \cdots \\
\alpha(\gamma(w)) & = & s_0 (e_1)^{2k} s_1 (e_2)^{2k} s_2 (e_3)^{2k} \cdots
\end{array}
$$

Therefore, since each element $e_i \in E(S)$ is an idempotent, $\textsc{eval}(w) = \alpha(\gamma(w))$.

Proof of Proposition 7.10: second item. Consider $K \in (C \circ \mathrm{SU}_k)(A)$, we have to build a new ω-language $H_K \in C(\mathbb{A}_\alpha)$ satisfying the following property:

$$
\text{For any well-formed } \omega\text{-word } w \in \mathbb{A}_\alpha^\omega, \quad w \in H_K \text{ if and only if } \gamma(w) \in K \tag{6}
$$

The argument is simpler that what we did for words since the definition of $C \circ \mathrm{SU}_k$ is less involved. By definition, there exists an SU_k-partition $\mathbf{P}$ of A^* and $L \in C(\mathbf{P} \times A)$ such that,

$$
K = \tau_{\mathbf{P}}^{-1}(L).
$$

The construction of H_K is based on the following lemma (which is adapted from Lemma 4.17 used in the case of finite words). Recall that any morphism $\beta : \mathbb{A}_\alpha^* \to (\mathbf{P} \times A)^*$ may be lifted as a map $\beta : \mathbb{A}_\alpha^\omega \to (\mathbf{P} \times A)^\omega$.

LEMMA 7.11. *There exists a morphism $\beta : \mathbb{A}_\alpha^* \to (\mathbf{P} \times A)^*$ such that for any well-formed ω-word $w \in \mathbb{A}_\alpha^\omega$, we have $\tau_{\mathbf{P}}(\gamma(w)) = \beta(w)$.*

The proof of Lemma 7.11 is identical to the one of Lemma 4.17. It is left to the reader. Let us use the lemma to construct H_K and finish the proof of Proposition 7.10. We have an ω-language $L \in C(\mathbf{P} \times A)$ such that $K = \tau_{\mathbf{P}}^{-1}(L)$. Let us define

$$
H_K = \beta^{-1}(L).
$$

Since C is closed inverse image, we obtain that $H_K \in C(\mathbb{A}_\alpha)$. We now prove that H_K satisfies (1) using Lemma 7.11. Given a well-formed ω-word $w \in \mathbb{A}_\alpha^\omega$, we have $w \in H_K$ if and only if $\beta(w) \in L$. The lemma then says that this is equivalent to $\tau_{\mathbf{P}}(\gamma(w)) \in L$, *i.e.*, to $\gamma(w) \in K$ by hypothesis on K.

7.5 From C-covering to $C \circ \mathrm{SU}$-covering

We now turn to the direction $(3) \Rightarrow (2)$ in Theorem 4.12. While the proof remains very similar to the one for the corresponding direction in Theorem 4.12, there is a significant technical difference. The argument is based on the following proposition adapted from Proposition 4.18. We let $n = 2^{3|S_+|+1}$ for the proof.

PROPOSITION 7.12. *There exists a map $\eta : A^\omega \to \mathbb{A}_\alpha^\omega$ satisfying the two following properties:*

(1) For any ω-language $L \subseteq A^\omega$ recognized by α, we have $L = \eta^{-1}(\mathrm{wF}_\alpha(L))$.

(2) For any ω-language $K \in C(\mathbb{A}_\alpha)$, the ω-language $\eta^{-1}(K)$ belongs to $(C \circ \mathrm{SU}_n)(A)$.

As before, one may show the direction $(3) \Rightarrow (2)$ in Theorem 7.6 from this proposition using an argument which is identical to the one used for proving the same direction in Theorem 4.12 from Proposition 4.18. Therefore, we leave it to the reader. We concentrate on proving Proposition 7.12.

Proof of Proposition 7.12: definition of η. We begin by defining $\eta : A^\omega \to \mathbb{A}_\alpha^\omega$. While similar, the definition is slightly different from the one we used when proving Proposition 4.18.

We first generalize the notion of k-*type* to ω-words. Given an ω-word w, a position x in w and a natural number $k \in \mathbb{N}$, a k-*type of x* is the following word of length at most k:

- If $x \leq k$, then the k-type of x is the prefix $w[1, x-1]$ of length $x-1$.
- If $x > k$, then the k-type of x is the infix $w[x-k, x-1]$ of length k.

For the construction of η, we fix $k = 2^{3|S_+|}$, so that $n = 2k$. Moreover, we choose an arbitrary order on the set of idempotents $E(S)$. We now generalize the notion of *distinguished* position to ω-words. The definition differs from the one we used for finite words. This change is needed to prove the first item in Proposition 7.12 (on the other hand, it is harmless for the second item: its proof is identical to the one for finite words). It is also the reason for using a larger constant k in this setting.

Consider an ω-word w and a position x in w. Moreover, let u be the k-type of x. We say that x is *distinguished* when there exists a nonempty suffix $v \in A^+$ of u such that $\alpha(v) \in E(S)$.

REMARK 7.13. *As for finite words, when x is distinguished, we have an idempotent $e \in E(S)$ such that $\alpha(u) \cdot e = \alpha(u)$ (namely, $e = \alpha(v)$ with v defined as above). However, in the case of ω-words, we have a stronger property: u has a suffix whose image under α is an idempotent. We need this to prove the first item in Proposition 7.12.*

We now generalize Fact 4.19: distinguished positions occur frequently in ω-words.

FACT 7.14. *Let $w \in A^\omega$, let $k = 2^{3|S_+|}$ and let $y \geq k - 1$ be some position of w. Then, there exists a distinguished position x in w such that $y - (k - 1) \leq x \leq y$.*

PROOF. It is known that every word $v \in A^+$ of length greater than $k = 2^{3|S_+|}$ contains an infix whose image under α is an idempotent (this is an immediate consequence of Simon's factorization forest theorem [8, 15, 40]). Hence, since the infix, $w[y-(k-1), y]$ has length k, the result follows. □

We may now define the map $\eta : A^\omega \to \mathbb{A}_\alpha^\omega$. Let $w \in A^\omega$. It is immediate from Fact 7.14 that w contains infinitely many distinguished positions. Let $x_0 < x_1 < x_2 < \cdots$ be these distinguished positions. For all $i \geq 0$, we let u_i be the k-type of x_i and $e_i \in E(S)$ be the smallest idempotent (according to the arbitrary order that we fixed on idempotents) such that u_i has a nonempty suffix whose image under α is e_i. We define $\eta(w) \in \mathbb{A}_\alpha^\omega$ as the ω-word:

$$\eta(w) = (\alpha(w_0), e_0) \cdot (e_0, \alpha(w_1), e_1) \cdot (e_1, \alpha(w_2), e_2) \cdot (e_2, \alpha(w_3), e_3) \cdots \in \mathbb{A}_\alpha^\omega$$

where $w_0 = w[0, x_0 - 1]$ and for all $i \geq 1$, $w_i = w[x_{i-1}, x_i - 1]$. Note that $\eta(w)$ is well-formed by definition. It remains to show that η satisfies the two items in Proposition 7.12.

Proof of Proposition 7.12: first item. This is where the technical differences with the proof of Proposition 4.18 occur. Consider a ω-language $L \subseteq A^\omega$ recognized by α and an ω-word $w \in A^\omega$. We have to show that $w \in L$ if and only if $\eta(w) \in \text{WF}_\alpha(L)$.

Since α recognizes L, we have $w \in L$ if and only if $\alpha(w) \in \alpha(L)$. Moreover, since $\eta(w)$ is well-formed, by definition, $\eta(w) \in \text{WF}_\alpha(L)$ if and only if $\text{EVAL}(\eta(w)) \in \alpha(L)$. Hence, it suffices to prove that $\text{EVAL}(\eta(w)) = \alpha(w)$. This requires more work than for words. By definition, we know that w may be decomposed as $w = w_0 w_1 w_2 \cdots$ and

$$\eta(w) = (\alpha(w_0), e_0) \cdot (e_0, \alpha(w_1), e_1) \cdot (e_1, \alpha(w_2), e_2) \cdot (e_2, \alpha(w_3), e_3) \cdots$$

such that for all $i \geq 0$, $w_0 \cdots w_i$ has a suffix of length at most k whose image under α is e_i. In particular, it is immediate by definition of EVAL that,

$$\text{EVAL}(\eta(w)) = \alpha(w_0)e_0\alpha(w_1)e_1\alpha(w_2)e_2\alpha(w_3)e_3 \cdots$$

Therefore, we need to show that,

$$\alpha(w_0)\alpha(w_1)\alpha(w_2)\alpha(w_3) \cdots = \alpha(w_0)e_0\alpha(w_1)e_1\alpha(w_2)e_2\alpha(w_3)e_3 \cdots$$

Using a standard Ramsey argument, we obtain an infinite sequence of indices $i_1, i_2, i_3, \ldots$ together with $s, t \in S$ and $f, g \in E(S)$ such that,

(1) $sf = s$ and $tg = t$.

(2) $\alpha(w_1) \cdots \alpha(w_{i_1}) = s$ and $\alpha(w_1)e_1 \cdots \alpha(w_{i_1})e_{i_1} = t$.

(3) For all $j > 1$, $\alpha(w_{i_{j-1}+1}) \cdots \alpha(w_{i_j}) = f$ and $\alpha(w_{i_{j-1}+1})e_{i_{j-1}+1} \cdots \alpha(w_{i_j})e_{i_j} = g$.

Therefore, $\alpha(w_0)\alpha(w_1)\alpha(w_2) \cdots = sf^\omega$ and $\alpha(w_0)e_0\alpha(w_1)e_1\alpha(w_2)e_2 \cdots = tg^\omega$. Hence, it suffices to prove that $sf^\omega = tg^\omega$. This is a consequence of the following lemma.

LEMMA 7.15. *We have $s = t$, $fg = f$ and $gf = g$.*

Before we prove the lemma, we show that $sf^\omega = tg^\omega$ and conclude the argument for the first item in Proposition 7.12. Using the lemma and the fact that $sf = s$, we obtain,

$$sf^\omega = s(fg)^\omega = sf(gf)^\omega = s(gf)^\omega = tg^\omega$$

It remains to prove Lemma 7.15. We prove the three equalities separately.

First Equality: $s = t$. By hypothesis, we know that for all $i \geq 0$, $w_0 \cdots w_i$ has a suffix of length at most k whose image under α is e_i. It follows that $\alpha(w_0 \cdots w_i) \cdot e_i = \alpha(w_0 \cdots w_i)$. Thus, it is immediate from a simple induction that indeed,

$$s = \alpha(w_1) \cdots \alpha(w_{i_1}) = \alpha(w_1)e_1 \cdots \alpha(w_{i_1})e_{i_1} = t.$$

Second Equality: $fg = f$. Let $j \geq 3$ be a large enough integer so that the word $w_{i_1+1}w_{i_1+2} \cdots w_{i_{j-1}}$ has length at least $k = 2^{3|S_+|}$. By definition and using the fact that f is idempotent, we have,

$$\begin{aligned}
\alpha(w_{i_1+1}w_{i_1+2} \cdots w_{i_{j-1}}) &= f, \\
\alpha(w_{i_{j-1}+1}) \cdots \alpha(w_{i_j}) &= f, \\
\alpha(w_{i_{j-1}+1})e_{i_{j-1}+1} \cdots \alpha(w_{i_j})e_{i_j} &= g.
\end{aligned}$$

Therefore it suffices to show that,

$$\alpha(w_{i_1+1} \cdots w_{i_{j-1}}) \cdot \alpha(w_{i_{j-1}+1}) \cdots \alpha(w_{i_j}) = \alpha(w_{i_1+1} \cdots w_{i_{j-1}}) \cdot \alpha(w_{i_{j-1}+1})e_{i_{j-1}+1} \cdots \alpha(w_{i_j})e_{i_j}.$$

By hypothesis, we know that for all $i \geq 0$, $w_0 \cdots w_i$ has a suffix **of length at most** k whose image under α is e_i. Thus, since $w_{i_1+1}w_{i_1+2} \cdots w_{i_{j-1}}$ has length at least k by hypothesis, it follows that for all $i \in \{i_{j-1} + 1, \ldots, i_j\}$, we have,

$$\alpha(w_{i_1+1} \cdots w_i) = \alpha(w_{i_1+1} \cdots w_i) \cdot e_i.$$

Hence, the result follows from a simple induction.

Third Equality: $gf = g$. Consider the word $w_{i_1+1} \cdots w_{i_2}$, which is mapped to f by α. By definition, for all $i_1 + 1 \leq j \leq i_2$, we have a suffix v_j of length at most k of $w_1 \cdots w_j$ such that $\alpha(v_j) = e_j$. We consider two sub-cases.

First assume that for all $i_1 + 1 \leq j \leq i_2$, the word v_j is a suffix of $w_{i_1+1} \cdots w_j$. In that case, it is immediate from as simple induction that we have,

$$f = \alpha(w_{i_1+1} \cdots w_{i_2}) = \alpha(w_{i_1+1})e_{i_1+1} \cdots \alpha(w_{i_2})e_{i_2} = g$$

Hence, we get $gf = g$. Otherwise, we consider the largest index j with $i_1 + 1 \leq j \leq i_2$ such that v_j is not a suffix of $w_{i_1+1} \cdots w_j$. Observe that,

a) Since j has been chosen to be maximal, we get from a simple induction that,

$$\alpha(w_{i_1+1} \cdots w_j)\alpha(w_{j+1})e_{j+1} \cdots \alpha(w_{i_2})e_{i_2} = \alpha(w_{i_1+1} \cdots w_{i_2}) = f. \tag{7}$$

b) Since v_j is by definition a suffix of $w_1 \cdots w_j$ but not of $w_{i_1+1} \cdots w_j$, it follows that $w_{i_1+1} \cdots w_j$ is a suffix of v_j.

c) Since $\alpha(v_j) = e_j$, we obtain from b) some $r \in S$ such that $r \cdot \alpha(w_{i_1+1} \cdots w_j) = e_j$.

Multiplying (7) by r on the left, we obtain therefore:

$$e_j\alpha(w_{j+1})e_{j+1}\cdots\alpha(w_{i_2})e_{i_2} = rf.$$

We may now multiply this equality on the left by $\alpha(w_{i_1+1})e_{i_1+1}\cdots e_{j-1}\alpha(w_j)$, which yields,

$$\alpha(w_{i_1+1})e_{i_1+1}\cdots\alpha(w_{i_2})e_{i_2} = \alpha(w_{i_1+1})e_{i_1+1}\cdots\alpha(w_j)rf,$$

that is,

$$g = \alpha(w_{i_1+1})e_{i_1+1}\cdots\alpha(w_j)rf.$$

In other words, we have found some element $r' \in S$ such that $g = r'f$. It follows that $gf = r'ff = r'f = g$, which concludes the proof.

Proof of Proposition 7.12: second item. Given an arbitrary ω-language $K \in C(\mathbb{A}_\alpha)$, we have to prove that $\eta^{-1}(K)$ belongs to $(C \circ \mathrm{SU}_{2k})(A)$ (recall that we fixed $n = 2k$). The proof of this item is essentially a simplified version of the corresponding argument for finite words (it is simpler since ω-words have no right "border").

Recall that $\sim_{2k}$ denotes the canonical equivalence associated to SU_{2k}. We let $\mathbf{P}$ be the partition of A^* into $\sim_{2k}$-classes. By definition of $C \circ \mathrm{SU}_{2k}$, it suffices to exhibit an ω-language $L \in C(\mathbf{P} \times A)$ such that $\eta^{-1}(K) = \tau_{\mathbf{P}}^{-1}(L)$. We use the following lemma (which is adapted from Lemma 4.21). Recall that we may lift a morphism $\beta : (\mathbf{P} \times A)^* \to \mathbb{A}_\alpha^*$ as a map $\beta : (\mathbf{P} \times A)^\omega \to \mathbb{A}_\alpha^\omega$.

LEMMA 7.16. *There exists a morphism* $\beta : (\mathbf{P} \times A)^* \to \mathbb{A}_\alpha^*$ *such that for any ω-word $w \in A^\omega$, we have* $\beta(\tau_{\mathbf{P}}(w)) = \eta(w)$.

The proof of Lemma 7.16 is identical to the one of Lemma 4.21 (using Fact 7.14 instead of Fact 4.19). We leave it to the reader. It remains to finish the proof of Proposition 7.12.

Let $\beta : (\mathbf{P} \times A)^* \to \mathbb{A}_\alpha^*$ be the morphism defined in Lemma 7.16. We claim that

$$\eta^{-1}(K) = \tau_{\mathbf{P}}^{-1}(\beta^{-1}(K)).$$

This will conclude the proof, since $\beta^{-1}(K)$ belongs to $C(\mathbf{P} \times A)$ by closure under inverse image. It remains to prove the claim. Let $w \in A^\omega$. By definition of β in Lemma 7.16, we have $w \in \eta^{-1}(K)$ if and only if $\beta(\tau_{\mathbf{P}}(w)) \in K$. This equivalent to $\tau_{\mathbf{P}}(w) \in \beta^{-1}(K)$. This exactly says that $w \in (\mathbf{P} \times A)^\omega \cap \tau_{\mathbf{P}}^{-1}(\beta^{-1}(K))$, as desired.

8 CONCLUSION

We presented generic reduction theorems for the SU-enrichment operation on classes of languages and ω-languages. Given any such class C satisfying appropriate closure properties, we reduce covering and separation for $C \circ \mathrm{SU}$ to the same problem for C.

These theorems have many applications: for most logical fragments, SU-enrichment is the language theoretic counterpart of a natural logical operation: if C is the class corresponding to some logical fragment, it is often the case that its SU-enrichment $C \circ \mathrm{SU}$ corresponds to the stronger fragment obtained by adding the predicates "+1", "min", "max" and "ε" to the signature. We showed this in the setting of finite words for the most prominent fragments of first-order logic, namely the two-variable fragment $\mathrm{FO}^2(<)$ and the levels $\Sigma_n(<)$ and $\mathcal{B}\Sigma_n(<)$ in the quantifier alternation hierarchy. Combined with our reduction theorem and already known results, this shows that covering and separation are decidable for $\mathrm{FO}^2(<, +1)$ and the levels $\Sigma_1(<, +1)$, $\mathcal{B}\Sigma_1(<, +1)$, $\Sigma_2(<, +1)$, $\mathcal{B}\Sigma_2(<, +1)$ and $\Sigma_3(<, +1)$. Note that several of these results were unknown, and that others have difficult combinatorial proofs only for membership algorithms (this is the case for $\mathcal{B}\Sigma_1(<, +1)$ [14] and for $\Sigma_2(<, +1)$ [12]).

An interesting follow-up to our work would be to obtain a similar reduction theorem for another natural operation: MOD-enrichment $C \mapsto C \circ \mathrm{MOD}$. Here, MOD stands for the class of modulo

languages. A language L belongs to MOD if and only if there exists a natural number $d \geq 1$ such that membership of a word w in L depends only $|w| \mod d$. This operation is important as it is the language theoretic counterpart of another natural logical operation. If C corresponds to some logical fragment, then $C \circ \text{MOD}$ corresponds to the stronger fragment obtained by adding the *modular predicates* to the signature. Essentially, they consist in unary predicates which can be used to test the number of a position modulo some constant.

REFERENCES

[1] Jorge Almeida. 1999. Some Algorithmic Problems for Pseudovarieties. *Publicationes Mathematicae Debrecen* 54 (1999), 531–552.

[2] Karl Auinger. 2010. On the Decidability of Membership in the Global of a Monoid Pseudovariety. *IJAC* 20, 2 (2010).

[3] Janusz A. Brzozowski and Robert Knast. 1978. The Dot-Depth Hierarchy of Star-Free Languages is Infinite. *J. Comput. System Sci.* 16, 1 (1978), 37–55.

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

[5] Julius Richard Büchi. 1960. Weak Second-Order Arithmetic and Finite Automata. *Mathematical Logic Quarterly* 6, 1-6 (1960), 66–92.

[6] Julius Richard Büchi. 1962. On a decision method in restricted second order arithmetic. In *Logic, Methodology and Philosophy of Science*. Stanford University Press.

[7] Olivier Carton. 2000. Wreath product and infinite words. *Journal of Pure and Applied Algebra* 153, 2 (2000), 129 – 150.

[8] Thomas Colcombet. 2010. Factorization Forests for Infinite Words and Applications to Countable Scattered Linear Orderings. *Theoritical Computer Science* 411, 4-5 (2010), 751–764.

[9] Wojciech Czerwiński, Wim Martens, and Tomáš Masopust. 2013. Efficient Separability of Regular Languages by Subsequences and Suffixes. In *Proceedings of the 40th International Colloquium on Automata, Languages, and Programming (ICALP'13)*. Springer-Verlag, Berlin, Heidelberg, 150–161.

[10] Samuel Eilenberg. 1976. *Automata, Languages, and Machines*. Vol. B. Academic Press, Inc., Orlando, FL, USA.

[11] Calvin C. Elgot. 1961. Decision Problems of Finite Automata Design and Related Arithmetics. *Trans. Amer. Math. Soc.* 98, 1 (1961), 21–51.

[12] Christian Glaßer and Heinz Schmitz. 2000. Languages of Dot-Depth 3/2. In *Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science ((STACS'00))*. Springer, Berlin, Heidelberg, 555–566.

[13] Christian Glaßer and Heinz Schmitz. 2007. Languages of Dot-Depth 3/2. *Theory of Computing Systems* 42, 2 (2007), 256–286.

[14] Robert Knast. 1983. A Semigroup Characterization of Dot-Depth One Languages. *RAIRO - Theoretical Informatics and Applications* 17, 4 (1983), 321–330.

[15] Manfred Kufleitner. 2008. The Height of Factorization Forests. In *Proceedings of the 33rd International Symposium on Mathematical Foundations of Computer Science (MFCS'08)*. Springer-Verlag, Berlin, Heidelberg, 443–454.

[16] Manfred Kufleitner and Alexander Lauser. 2012. Around dot-Depth 1. *Int. J. Found. Comp. Sci.* 23, 6 (2012).

[17] Robert McNaughton. 1974. Algebraic decision procedures for local testability. *Mathematical Systems Theory* 8, 1 (1974), 60–76.

[18] Robert McNaughton and Seymour A. Papert. 1971. *Counter-Free Automata*. MIT Press.

[19] Dominique Perrin. 1984. Recent Results on Automata and Infinite Words. In *Proceedings of the 9th International Symposium on Mathematical Foundations of Computer Science (MFCS'84)*. Springer-Verlag, Berlin, Heidelberg, 134–148.

[20] Dominique Perrin and Jean-Éric Pin. 2004. *Infinite Words*. Elsevier.

[21] Théo Pierron, Thomas Place, and Marc Zeitoun. 2016. Quantifier Alternation for Infinite Words. In *Foundations of Software Science and Computation Structures - 19th International Conference, FOSSACS 2016*. 234–251.

[22] Jean-Éric Pin and Pascal Weil. 1995. Polynomial Closure and Unambiguous Product. In *Proceedings of the 22nd International Colloquium on Automata, Languages and Programming (ICALP'95)*. Springer-Verlag, Berlin, Heidelberg, 348–359.

[23] Jean-Éric Pin. 1995. A variety theorem without complementation. *Russian Mathematics (Izvestija vuzov.Matematika)* 39 (1995), 80–90.

[24] Jean-Éric Pin and Pascal Weil. 1997. Polynomial Closure and Unambiguous Product. *Theory of Computing Systems* 30, 4 (1997), 383–422.

[25] Jean-Eric Pin and Pascal Weil. 2002. The wreath product principle for ordered semigroups. *Communications in Algebra* 30 (2002), 5677–5713.

[26] Thomas Place. 2015. Separating Regular Languages with Two Quantifiers Alternations. In *Proceedings of the 30th Annual ACM/IEEE Symposium on Logic in Computer Science, (LICS'15)*. IEEE Computer Society, 202–213.

[27] Thomas Place. 2017. Separating Regular Languages with Two Quantifiers Alternations. (2017). Unpublished, see https://arxiv.org/abs/1707.03295 for a preliminary version.

[28] Thomas Place, Lorijn van Rooijen, and Marc Zeitoun. 2013. Separating Regular Languages by Piecewise Testable and Unambiguous Languages. In *Proceedings of the 38th International Symposium on Mathematical Foundations of Computer Science (MFCS'13)*. Springer-Verlag, Berlin, Heidelberg, 729–740.

[29] Thomas Place and Marc Zeitoun. 2014. Going Higher in the First-Order Quantifier Alternation Hierarchy on Words. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP'14)*. Springer-Verlag, Berlin, Heidelberg, 342–353.

[30] Thomas Place and Marc Zeitoun. 2014. Separating Regular Languages with First-order Logic. In *Proceedings of the Joint Meeting of the 23rd EACSL Annual Conference on Computer Science Logic (CSL'14) and the 29th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'14)*. ACM, New York, NY, USA, 75:1–75:10.

[31] Thomas Place and Marc Zeitoun. 2015. Separation and the Successor Relation. In *32nd International Symposium on Theoretical Aspects of Computer Science (STACS'15)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 662–675.

[32] Thomas Place and Marc Zeitoun. 2016. The Covering Problem: A Unified Approach for Investigating the Expressive Power of Logics. In *Proceedings of the 41st International Symposium on Mathematical Foundations of Computer Science (MFCS'16)*. 77:1–77:15.

[33] Thomas Place and Marc Zeitoun. 2016. Separating Regular Languages with First-Order Logic. *Logical Methods in Computer Science* 12, 1 (2016).

[34] Thomas Place and Marc Zeitoun. 2017. The Covering Problem. (2017). Unpublished, a preliminary version can be found at https://arxiv.org/abs/1707.03370.

[35] Thomas Place and Marc Zeitoun. 2017. Going Higher in First-Order Quantifier Alternation Hierarchies on Words. (2017). arXiv, https://arxiv.org/pdf/1707.05696.pdf .

[36] Thomas Place and Marc Zeitoun. 2017. Separation for Dot-depth Two. In *Proceedings of the 32th Annual ACM/IEEE Symposium on Logic in Computer Science, (LICS'17)*. IEEE Computer Society, 202–213.

[37] Michael O. Rabin. 1969. Decidability of Second-Order Theories And Automata on Infinite Trees. *Trans. Amer. Math. Soc.* 141, 1-35 (1969), 4.

[38] Marcel Paul Schützenberger. 1965. On Finite Monoids Having Only Trivial Subgroups. *Information and Control* 8, 2 (1965), 190–194.

[39] Imre Simon. 1975. Piecewise Testable Events. In *Proceedings of the 2nd GI Conference on Automata Theory and Formal Languages*. Springer-Verlag, Berlin, Heidelberg, 214–222.

[40] Imre Simon. 1990. Factorization Forests of Finite Height. *Theoritical Computer Science* 72, 1 (1990), 65–94.

[41] Benjamin Steinberg. 2001. A delay theorem for pointlikes. *Semigroup Forum* 63, 3 (2001), 281–304.

[42] Howard Straubing. 1985. Finite Semigroup Varieties of the Form V ∗ D. *Journal of Pure and Applied Algebra* 36 (1985), 53–94.

[43] James W. Thatcher and Jesse B. Wright. 1968. Generalized Finite Automata Theory with an Application to a Decision Problem of Second-Order Logic. *Mathematical Systems Theory* 2, 1 (1968), 57–81.

[44] Denis Thérien and Thomas Wilke. 1998. Over Words, Two Variables Are As Powerful As One Quantifier Alternation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC'98)*. ACM, New York, NY, USA, 234–240.

[45] Boris A. Trakhtenbrot. 1961. Finite Automata and Logic of Monadic Predicates. *Doklady Akademii Nauk SSSR* 149 (1961), 326–329. In Russian.

[46] Thomas Wilke. 1991. An Eilenberg Theorem for Infinity-Languages. In *Proceedings of the 18th International Colloquium on Automata, Languages and Programming (ICALP'91)*. Springer-Verlag, Berlin, Heidelberg, 588–599.

[47] Yechezkel Zalcstein. 1972. Locally testable languages. *J. Comput. System Sci.* 6, 2 (1972), 151–167.