HAL
open science

Positive First-Order Logic on Words and Graphs

Denis Kuperberg

To cite this version:

Denis Kuperberg. Positive First-Order Logic on Words and Graphs. Logical Methods in Computer Science, 2023, 19 (3), pp.7:1 - 7:35. 10.46298/LMCS-19(3:7)2023 . hal-03865495v2

HAL Id: hal-03865495

https://hal.science/hal-03865495v2

Submitted on 17 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Logical Methods in Computer Science
Volume 19, Issue 3, 2023, pp. 7:1–7:35
https://lmcs.episciences.org/

Submitted Feb. 15, 2022
Published Jul. 25, 2023

POSITIVE FIRST-ORDER LOGIC ON WORDS AND GRAPHS *

DENIS KUPERBERG

CNRS, LIP, ENS Lyon
e-mail address: denis.kuperberg@ens-lyon.fr

Abstract. We study FO^+, a fragment of first-order logic on finite words, where monadic predicates can only appear positively. We show that there is an FO-definable language that is monotone in monadic predicates but not definable in FO^+. This provides a simple proof that Lyndon's preservation theorem fails on finite structures. We lift this example language to finite graphs, thereby providing a new result of independent interest for FO-definable graph classes: negation might be needed even when the class is closed under addition of edges. We finally show that the problem of whether a given regular language of finite words is definable in FO^+ is undecidable.

1. Introduction

Preservation theorems in first-order logic (FO) establish a link between semantic and syntactic properties [AG97, Ros08]. We will be particularly interested here in Lyndon's theorem [Lyn59], which states that if a first-order formula is monotone in a predicate P (semantic property), then it is equivalent to a formula that is positive in P (syntactic property). Recall that "monotone in P" means that the formula stays true when tuples are added to P, while "positive in P" means that P does not appear under a negation in the formula.

As it is often the case with preservation theorems, Lyndon's Theorem may not hold when restricting the class of structures considered. Whether Lyndon's Theorem is true when restricted to finite structures was an open problem for 28 years. It was finally shown to fail on finite structures in [AG87] with a very difficult proof, using a large array of techniques from different fields of mathematics such as probability theory, topology, lattice theory, and analytic number theory. A simpler but still quite intricate proof of this fact was later given by [Sto95], using Ehrenfeucht-Fraïssé games on grid-like structures equipped with two binary predicates. This construction was slightly modified in [Ros95] to treat a signature monotone in every relation symbol.

The goal of this paper is to further restrict the class of structures under consideration, starting by allowing only finite words. This will allow us to obtain in turn a better understanding of the problem for finite graphs and general finite structures. We will therefore work in most of this paper with the particular signature associated with finite words: one binary predicate (the total order on positions in the word), and a finite set of monadic predicates (encoding the alphabet). We will call FO^+ the fragment of first-order

* This paper is an extended version of [Kup21], the main new content is the section about graphs.

DOI:10.46298/LMCS-19(3:7)2023
© D. Kuperberg
Creative Commons

logic on these models, that is syntactically positive in the monadic predicates. Or more simply: FO^+ is the negation-free fragment of FO on finite words.

Our purpose is twofold:

- Find out whether Lyndon's Theorem holds on finite words, and investigate the relation of this framework with the more general case of finite structures, as well as finite graphs.
- From the point of view of language theory: study the natural fragment of FO^+-definable languages, in particular given a regular language, can we decide whether it is FO^+-definable?

Recall that FO on words is a well-studied logic defining a proper fragment of regular languages. This fragment has many equivalent characterizations: definable by star-free expressions, aperiodic monoids, LTL formulas,... [DG08, Sch65, Kam68, MP71]. In particular, given a regular language, it is decidable whether it is FO-definable.

Contributions. We define a semantic notion of monotone language on alphabets equipped with a partial order: the language is required to be closed under replacement of a letter by a bigger one. This generalizes the monotonicity condition on monadic predicates in the sense of Lyndon. The negation-free logic FO^+ can only define monotone languages, and can be seen as a fragment of the standard FO logic on words, in the context of ordered alphabets.

Answering our first objective, we show that Lyndon's Theorem fails on finite words, by building a regular language that is monotone and FO-definable, but not FO^+-definable. This proof uses a variant of Ehrenfeucht-Fraïssé games that characterizes FO^+-definability, introduced in [Sto95], and instantiated here on finite words. As a corollary, using suitable axiomatizations of finite words, we obtain the failure of Lyndon's theorem on finite structures, in a much simpler way than in [AG87, Sto95]. We show that this can be done regardless of the precise version of Lyndon's preservation theorem that is considered: either monotonicity is with respect to one predicate, or an arbitrary subset of them, or all of them. These different versions are discussed in [Lyn59, AG87, Sto95]. The version where all predicates are monotone, can be reformulated as closure under surjective homomorphisms [Lyn59].

We also show that our counter-example language can be lifted to finite graphs, both directed and undirected, using a suitable encoding. We thereby show that there exists an FO-definable class of graphs, closed under edge addition, but such that any FO formula defining this class must use some edge predicate under a negation. To our knowledge, this is a new result, and we believe it is of independent interest, as it provides a surprising answer to a natural question about graphs.

Finally, answering our second objective, we show that FO^+-definability is undecidable for regular languages. This result is obtained using a reduction from the Turing Machine Mortality problem [Hoo66]. To our knowledge, this is the first example of a natural[1] class of regular languages for which membership is undecidable.

Although we work in a specialized framework requiring a partial order on the alphabet, we believe that this is not an artificial construct, as it occurs naturally in several settings. On one hand, powerset alphabets – i.e. letters are sets of atomic predicates, and are naturally ordered by inclusion – are standard in verification and model theory. It is of interest to remark that all the results of this paper are true in the particular case of powerset alphabets,

[1]The concept of "natural" class is of course a bit informal here, but for instance we can think of it as classes inductively defined via a syntax. More generally, any class of regular languages not purposely defined to have an undecidable membership problem could be considered natural in this context.

and that this allows to obtain the failure of Lyndon's theorem on general finite structures. On the other hand, ordered letters can be used as an abstraction for factors of the input word that are naturally equipped with a partial order, for instance in quantitative generalizations of regular languages such as regular cost functions [Col12].

Related works.

Monotone complexity:

Positive fragments of first-order logic play a prominent role in complexity theory. Indeed, an active research program consists in studying positive fragments of complexity classes. This includes for instance trying to lift equivalent characterizations of a class to their positive versions, or investigating whether a semantic and a syntactic definition of the positive variant of a class are equivalent. See [GS92] for an introduction to monotone complexity, and [LSS96, Ste94] for examples of characterizations of the positive versions of the classes P and NP, in particular through extensions of first-order logic. The aforementioned paper [AG87], which was the first to show the failure of Lyndon's theorem on finite structures, does so by reproving in particular an important result on monotone circuit complexity first proved in [FSS81]: Monotone-$\mathrm{AC}^0 \neq$ Monotone $\cap\, \mathrm{AC}^0$.

Membership in subclasses of regular languages:

Related to our undecidability result, we can mention that there are syntactically defined classes of regular languages for which decidability of membership is an open problem. Such classes, also related to FO fragments, are the ones defined via quantifier-alternation: given a regular language, is it definable with an FO formula having at most k quantifier alternations? Recent works obtained decidability results for this question, but only for the first 3 levels of the quantifier alternation hierarchy [PZ19]. For higher levels, the problem remains open. Let us also mention the generalized star-height problem [PST89]: can a given regular language be defined in an extended regular expression (with complement allowed) with no nesting of Kleene star? In this case it is not even known whether all regular languages can be defined in this way.

Quantitative extensions:

First-order logic on words has been extended to quantitative settings, which naturally yields a negation-free syntax, because complementation becomes problematic in those settings. This is the case in the theory of regular cost functions [Col12, KVB12], and in other quantitative extensions concerned with boundedness properties, such as MSO+U [Boj04] or Magnitude MSO [Col13]. We hope that the present work can shed a light on these extensions as well.

Notations and prerequisites. If $i, j \in \mathbb{N}$, we note $[i, j]$ the set $\{i, i+1, \ldots, j\}$. If X is a set, we note $|X|$ its cardinal, and $\mathcal{P}(X)$ its powerset, i.e. the set of subsets of X. We will note A a finite alphabet throughout the paper. The set of finite words on A is A^*. The length of $u \in A^*$ is denoted $|u|$. If $L \subseteq A^*$ is a language, we will note $\overline{L}$ its complement. We will note $dom(u) = [0, |u|-1]$ the set of positions of a word u. If u is a word and $i \in dom(u)$, we will note $u[i]$ the letter at position i, and $u[..i]$ the prefix of u up to position i included. Similarly, $u[i..j]$ is the infix of u from position i to j included and $u[i..]$ is the suffix of u starting in position i included.

We will assume that the reader is familiar with the notion of regular languages of finite words, and with some ways to define such languages: finite automata (DFA for deterministic

and NFA for non-deterministic), finite monoids, and first-order logic. See e.g. [DG08] for an introduction to all the needed material.

2. Monotonicity on words

2.1. **Ordered alphabet.** In this paper we will consider that the finite alphabet A is equipped with a partial order $\leq_A$. This partial order is naturally extended to words componentwise: $a_1a_2\dots a_n \leq_A b_1b_2\dots b_m$ if $n = m$ and for all $i \in [1, n]$ we have $a_i \leq_A b_i$.

A special case that will be of interest here is when the alphabet is built as the powerset of a set P of *predicates*, i.e. $A = \mathcal{P}(P)$, and the order $\leq_A$ is inclusion. We will call this a *powerset alphabet*.

Taking $A = \mathcal{P}(P)$ is standard in settings such as verification and model theory, where several predicates can be considered independently of each other in some position.

Powerset alphabets constitute a particular case of ordered alphabets. The results obtained in this paper are valid for both the powerset case and the general case. Due to the nature of the results (existence of a counter-example and undecidability result), it is enough to show them in the particular case of powerset alphabets to cover both cases. Moreover, the powerset alphabet case allows us to directly establish a link with Lyndon's theorem, which is stated in the framework of model theory. For these reasons, we will keep the more general notion of ordered alphabet for generic definitions, but we will prove our main results on powerset alphabets in order to directly obtain the stronger version of these results.

2.2. **Monotone languages.** We fix A a finite ordered alphabet.

Definition 2.1. We say that a language $L \subseteq A^*$ is *monotone* if for all $u \leq_A v$, if $u \in L$ then $v \in L$.

Example 2.2. Let $A = \{a, b\}$ with $a \leq_A b$. Then A^*bA^* is monotone but its complement a^* is not monotone.

Definition 2.3. Let $L \subseteq A^*$, the *monotone closure* of L is the language $L^\uparrow = \{v \in A^* \mid \exists u \in L, u \leq_A v\}$. It is the smallest monotone language containing L.

In particular, if $a \in A$, we will note $a^\uparrow$ the set $\{b \in A \mid a \leq_A b\}$.

Lemma 2.4. *Given an NFA $\mathcal{B}$, we can compute in time $O(|\mathcal{B}| \cdot |A|)$ an NFA $\mathcal{B}^\uparrow$ for the monotone closure of $L(\mathcal{B})$.*

Proof. We build an NFA $\mathcal{B}^\uparrow$ from $\mathcal{B}$, by replacing every transition $p \xrightarrow{a} q$ of $\mathcal{B}$ by $p \xrightarrow{a^\uparrow} q$. We use here the standard convention where a transition $p \xrightarrow{X} q$ with $X \subseteq A$ stands for a set of transitions $\{p \xrightarrow{b} q \mid b \in X\}$. It is straightforward to verify that $\mathcal{B}^\uparrow$ is an NFA for $L^\uparrow$: any run of $\mathcal{B}^\uparrow$ on some word v can be mapped to a run of $\mathcal{B}$ on some $u \leq_A v$. □

Theorem 2.5. *Given a regular language $L \subseteq A^*$, it is decidable whether L is monotone. The problem is in* P *if L is given by a DFA and* Pspace*-complete if L is given by an NFA, on any alphabet with non-trivial order.*

Proof. Notice that if $\mathcal{B}$ is an NFA, $L(\mathcal{B})$ is monotone if and only if $L(\mathcal{B}^{\uparrow}) \subseteq L(\mathcal{B})$. This shows that the problem is in PSPACE in general, and that it is in P when $\mathcal{B}$ is a DFA, since it reduces to checking emptiness of the intersection between $\mathcal{B}^{\uparrow}$ and the complement of $\mathcal{B}$. We show that the general problem is PSPACE-hard by reducing from NFA universality. Let $\mathcal{B}$ be an NFA on a two-letter alphabet $A = \{a, b\}$. We build an NFA $\mathcal{C}$ of size polynomial in the size of $\mathcal{B}$ and recognizing $aA^* + bL(\mathcal{B})$, using standard NFA constructions. We now consider the monotonicity of $\mathcal{C}$ according to the alphabet order $a \leq_A b$. If $L(\mathcal{C})$ is monotone, since for all $u \in A^*$ we have $au \in L(\mathcal{C})$, we obtain $bu \in L(\mathcal{C})$ as well, so $u \in L(\mathcal{B})$. So $L(\mathcal{C})$ monotone implies $\mathcal{B}$ universal. Conversely, if $\mathcal{B}$ is universal, then $L(\mathcal{C}) = A^*$ hence $L(\mathcal{C})$ is monotone. We have that $L(\mathcal{B}) = A^*$ if and only if $L(\mathcal{C})$ is monotone, thereby completing the PSPACE-hardness reduction. This means that the monotonicity problem is PSPACE-complete as soon as there are two comparable letters $a \leq_A b$ in the alphabet. Otherwise the problem is trivial, as any language is monotone on a trivially ordered alphabet. □

3. Positive first-order logic

3.1. **Syntax and semantics.** The main idea of positive FO, that we will note FO^+, is to guarantee via a syntactic restriction that it only defines monotone languages.

Notice that since monotone languages are not closed under complement (see Example 2.2), we cannot allow negation in the syntax of FO^+. This means we have to add dual versions of classical operators of first-order logic.

This naturally yields the following syntax for FO^+:

$$\varphi, \psi := a^{\uparrow}(x) \mid x \leq y \mid x < y \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \exists x.\varphi \mid \forall x.\varphi$$

As usual, variables $x, y, \ldots$ range over the positions of the input word. The semantics is the same as classical FO on words, with the notable exception that $a^{\uparrow}(x)$ is true if and only if x is labelled by some $b \in a^{\uparrow}$. Unlike classical FO, it is not possible to require that a position is labelled by a specific letter a, except when $a^{\uparrow} = \{a\}$. This is necessary to guarantee that only monotone languages can be defined.

Formal semantics of FO^+

If φ is a formula with free variables $\mathrm{FV}(\varphi)$, its semantics is a set $[\![\varphi]\!]$ of pairs of the form (u, α), where $u \in A^*$ and $\alpha : \mathrm{FV}(\varphi) \to dom(u)$ a valuation for the free variables. We write indistinctively $u, \alpha \models \varphi$ or $(u, \alpha) \in [\![\varphi]\!]$, to signify that (u, α) satisfies φ. If $\mathrm{FV}(\varphi) = \emptyset$, we can simply write $u \models \varphi$ instead of $(u, \emptyset) \models \varphi$. In this case, the language recognized by φ is $[\![\varphi]\!] = \{u \in A^* \mid u \models \varphi\}$.

We define $[\![\varphi]\!]$ by induction on φ.

- $u, \alpha \models a^{\uparrow}(x)$ if $a \leq_A u[\alpha(x)]$.
- $u, \alpha \models x \leq y$ if $\alpha(x) \leq \alpha(y)$.
- $u, \alpha \models x < y$ if $\alpha(x) < \alpha(y)$.
- $u, \alpha \models \varphi \vee \psi$ if $u, \alpha \models \varphi$ or $u, \alpha \models \psi$.
- $u, \alpha \models \varphi \wedge \psi$ if $u, \alpha \models \varphi$ and $u, \alpha \models \psi$.
- $u, \alpha \models \exists x.\varphi$ if there exists $i \in dom(u)$ such that $u, \alpha[x \mapsto i] \models \varphi$.
- $u, \alpha \models \forall x.\varphi$ if for all $i \in dom(u)$, we have $u, \alpha[x \mapsto i] \models \varphi$.

Here the valuation $\alpha[x \mapsto i]$ maps y to $\begin{cases} i & \text{if } y = x \\ \alpha(y) & \text{if } y \neq x \end{cases}$.

Example 3.1. On alphabet $A = \{a, b, c\}$ with $a \leq_A b$.
- $\forall x.a^{\uparrow}(x)$ recognizes $\{a, b\}^*$.
- $\exists x.b^{\uparrow}(x)$ recognizes A^*bA^*.

Remark 3.2. In the powerset alphabet framework where $A = \mathcal{P}(P)$, we can naturally view FO^+ as the negation-free fragment of first-order logic, by having atomic predicates $a^{\uparrow}(x)$ range directly over P instead of $A = \mathcal{P}(P)$. We can then drop the $a^{\uparrow}$ notation, as predicates from P are considered independently of each other. This way, $p(x)$ will be true if and only if the letter $S \in A$ labelling x contains p. A letter predicate $S^{\uparrow}(x)$ in the former syntax can then be expressed by $\bigwedge_{p \in S} p(x)$, so FO^+ based on predicates from P is indeed equivalent to FO^+ based on A. We will take this convention when working on powerset alphabets.

Example 3.3. Let $A = \mathcal{P}(P)$ with $P = \{a, b\}$. The formula $\exists x, y.\ x \leq y \wedge a(x) \wedge b(y)$ recognizes $A^*\{a, b\}A^* + A^*\{a\}A^*\{b\}A^*$.

3.2. Properties of FO^+.

Lemma 3.4. *Assume the order on A is trivial, i.e. no two distinct letters are comparable. Then all languages are monotone, and any FO-definable language is FO^+-definable.*

Proof. The fact that all languages are monotone in this case follows from the fact that for two words u, v we have $u \leq_A v$ if and only if $u = v$.

If L is definable by an FO formula φ, we can build an FO^+ formula ψ from φ by pushing negations to the leaves using the usual rewritings such as $\neg(\varphi \wedge \psi) = \neg\varphi \vee \neg\psi$ and $\neg(\exists x.\varphi) = \forall x.\neg\varphi$. For all letter $a \in A$ and variable x, we then replace all occurrences of $\neg a(x)$ by $\bigvee_{b \neq a} b(x)$. Finally, the negation of $x \leq y$ (resp. $x < y$) can be written $y < x$ (resp. $y \leq x$). □

Lemma 3.5. *The logic FO^+ can only define monotone languages.*

Proof. This is done by induction on the FO^+ formula φ, where the induction property is strengthened to include possible free variables: for all $(u, \alpha) \in [\![\varphi]\!]$ and $v \geq_A u$, we have $(v, \alpha) \in [\![\varphi]\!]$.

Base cases:

Let $(u, \alpha) \in [\![a^{\uparrow}(x)]\!]$ and $v \geq_A u$, we have $v[\alpha(x)] \geq_A u[\alpha(x)] \geq_A a$, so $(v, \alpha) \in [\![a^{\uparrow}(x)]\!]$.

Let $(u, \alpha) \in [\![x \leq y]\!]$ and $v \geq_A u$. We have $\alpha(x) \leq \alpha(y)$ so $(v, \alpha) \in [\![x \leq y]\!]$. The argument for $<$ instead of $\leq$ is identical.

Induction cases:

Let $(u, \alpha) \in [\![\varphi \vee \psi]\!]$ and $v \geq_A u$. We have $(u, \alpha) \in [\![\varphi]\!]$ or $(u, \alpha) \in [\![\psi]\!]$. Therefore, by induction hypothesis, $(v, \alpha) \in [\![\varphi]\!]$ or $(v, \alpha) \in [\![\psi]\!]$, hence $(v, \alpha) \in [\![\varphi \vee \psi]\!]$. The argument for $\varphi \vee \psi$ is identical.

Let $(u, \alpha) \in [\![\exists x.\varphi]\!]$ and $v \geq_A u$. There exists $i \in dom(u)$ such that $(u, \alpha[x \mapsto i]) \in [\![\varphi]\!]$. By induction hypothesis, $(v, \alpha[x \mapsto i]) \in [\![\varphi]\!]$. Hence, $(v, \alpha) \in [\![\exists x.\varphi]\!]$. The argument for $\forall$ is identical. □

It is natural to ask whether the converse of Lemma 3.5 holds: if a language is FO-definable and monotone, then is it necessarily FO^+-definable? This will be the purpose of Section 4.

3.3. **Ordered Ehrenfeucht-Fraïssé games.** We will explain here how FO^+-definability can be captured by an ordered variant of Ehrenfeucht-Fraïssé games, that we will call EF^+-games.

This notion was defined in [Sto95] for general structures, we will instantiate it here on finite words.

We define the n-round EF^+-game on two words $u, v \in A^*$, noted $\mathrm{EF}_n^+(u, v)$. This game is played between two players, Spoiler and Duplicator.

If $k \in \mathbb{N}$, a *k-position* of the game is of the form (u, α, v, β), where $\alpha : [1, k] \to dom(u)$ and $\beta : [1, k] \to dom(v)$ are valuations for k variables in u and v respectively. We can think of α and β as giving the position of k previously placed tokens in u and v.

A k-position (u, α, v, β) is *valid* if for all $i \in [1, k]$, we have $u[\alpha(i)] \leq_A v[\beta(i)]$, and for all $i, j \in [1, k]$, $\alpha(i) \leq \alpha(j)$ if and only if $\beta(i) \leq \beta(j)$.

Notice the difference with usual EF-games: here we do not ask that tokens placed in the same round have same label, but that the label in u is $\leq_A$-smaller than the label in v. This feature is intended to capture FO^+ instead of FO.

The game starts from the 0-position $(u, \emptyset, v, \emptyset)$.

At each round, starting from a k-position (u, α, v, β), the game is played as follows. If $k = n$, then Duplicator wins. Otherwise, Spoiler chooses a position in one of the two words, and places token number $k + 1$ on it. Duplicator answers by placing token number $k + 1$ on a position of the other word. Let us call α' and β' the extensions of α and β with these new tokens. If (u, α', v, β') is not a valid $(k + 1)$-position, then Spoiler immediately wins the game, otherwise, the game moves to the next round with $(k + 1)$-position (u, α', v, β').

We will note $u \preceq_n v$ when Duplicator has a winning strategy in $\mathrm{EF}_n^+(u, v)$.

Definition 3.6. The quantifier rank of a formula φ, noted $\mathrm{qr}(\varphi)$ is its maximal number of nested quantifiers. It can be defined by induction in the following way: if φ is atomic then $\mathrm{qr}(\varphi) = 0$, otherwise, $\mathrm{qr}(\varphi \wedge \psi) = \mathrm{qr}(\varphi \vee \psi) = \max(\mathrm{qr}(\varphi), \mathrm{qr}(\psi))$ and $\mathrm{qr}(\exists x.\varphi) = \mathrm{qr}(\forall x.\varphi) = \mathrm{qr}(\varphi) + 1$.

The following Theorem shows the link between the n-round EF^+ game and formulas of rank at most n.

Theorem 3.7 [Sto95, Thm 2.4]**.** *We have $u \preceq_n v$ if and only if for all formulas φ of* FO^+ *with $\mathrm{qr}(\varphi) \leq n$, we have $(u \models \varphi) \Rightarrow (v \models \varphi)$.*

Since the proof of Theorem 3.7 does not appear in [Sto95], we will prove it in a general setting in Section 5.2.

Let us now see how we can use EF^+ games to characterize FO^+-definability.

Corollary 3.8. *A language L is not* FO^+*-definable if and only if for all $n \in \mathbb{N}$, there exists $(u, v) \in L \times \overline{L}$ such that $u \preceq_n v$.*

Proof. $\Leftarrow$: Let $n \in \mathbb{N}$, there exists $(u, v) \in L \times \overline{L}$ such that $u \preceq_n v$. By Theorem 3.7, any formula of quantifier rank n accepting u must accept v, so no formula of quantifier rank n recognizes L. This is true for all $n \in \mathbb{N}$, so L is not FO^+-definable.

$\Rightarrow$ (contrapositive): Assume there exists $n \in \mathbb{N}$ such that for all $(u, v) \in L \times \overline{L}$, $u \not\preceq_n v$. By Theorem 3.7, this means that for all $(u, v) \in L \times \overline{L}$, there exists a formula $\varphi_{u,v}$ of quantifier rank n accepting u but not v. Since there are finitely many FO^+ formulas of rank n up to logical equivalence [Lib04, Lem 3.13], the set of formulas $F = \{\varphi_{u,v} \mid (u, v) \in L \times \overline{L}\}$ can be chosen finite. We define $\psi = \bigvee_{u \in L} \bigwedge_{v \notin L} \varphi_{u,v}$, where the conjunctions and disjunction

are finite since F is finite. For all $u \in L$, $u \models \bigwedge_{v \notin L} \varphi_{u,v}$ hence $u \models \psi$, and conversely, a word satisfying ψ must satisfy some $\bigwedge_{v \notin L} \varphi_{u,v}$, so it cannot be in $\overline{L}$. □

4. A counter-example language

4.1. **The language K.** We will now answer the natural question posed in Section 3.2: is any FO-definable monotone language (on any ordered alphabet) also FO^+-definable?

This section is dedicated to the proof of the following Theorem:

Theorem 4.1. *There is an FO-definable monotone language K on a powerset alphabet that is not* FO^+*-definable.*

Let $P = \{a, b, c\}$ and $A = \mathcal{P}(P)$, ordered by inclusion.

We will note $\binom{a}{b}, \binom{b}{c}, \binom{c}{a}$ for the letters $\{a, b\}, \{b, c\}, \{a, c\}$ respectively, and $\top$ for $\{a, b, c\}$. If $x \in P$ we will often note x instead of $\{x\}$ to lighten notations.

Definition 4.2. We now define the desired language by:

$$K := (a^{\uparrow} b^{\uparrow} c^{\uparrow})^* + A^* \top A^*.$$

We claim that K satisfies the requirements of Theorem 4.1.

Notice that the second disjunct $A^* \top A^*$ could be omitted if we were to consider only the alphabet $A \setminus \{\top\}$. When sticking with a powerset alphabet, this disjunct is necessary to obtain an FO-definable language. Indeed, if we just define $K_0 = (a^{\uparrow} b^{\uparrow} c^{\uparrow})^*$ on alphabet A, we have $K_0 \cap (\top^*) = (\top\top\top)^*$. Since $\top^*$ is FO-definable but $(\top\top\top)^*$ is not (see [DG08]), and FO-definable languages are closed under intersection, we have that K_0 is not FO-definable.

4.2. **FO-definability of K.**

Lemma 4.3. *K is monotone and FO-definable.*

Proof. The fact that K is monotone is straightforward from its definition, as the union of two monotone languages.

We will show that K is FO-definable in three different ways: using its minimal DFA, its syntactic monoid, and finally describing how an FO formula recognizing it can be defined. This gives several complementary points of view, which can all be helpful for a deep understanding of this counter-example language.

Let us start with the automaton approach. We use the classical characterizations of first-order definable languages [DG08] by verifying that the minimal DFA $\mathcal{A}$ of K is counter-free. That is, no word induces a non-trivial cycle in $\mathcal{A}$.

The minimal DFA $\mathcal{A}$ recognizing K is depicted in Figure 1. We note $\neg a = \{\emptyset, \{b\}, \{c\}, \{b, c\}\}$ the sub-alphabet of A of letters not containing a, similarly for $\neg b$ and $\neg c$. The edges going to rejecting state $\perp$ are grayed and dashed, and the ones going to accepting sink state $q_\top$ are grayed, for readability. We also note $a' = a^{\uparrow} \setminus \{\top\} = \{\{a\}, \binom{a}{b}, \binom{c}{a}\}$, and similarly for b', c'.

To show that K is FO-definable, it suffices to show that $\mathcal{A}$ is counter-free, i.e. that there is no word $u \in A^*$, distinct states p, q of $\mathcal{A}$, and integer k, such that $p \xrightarrow{u} q$ and $q \xrightarrow{u^k} p$. Assume for contradiction that such u, p, q, k exist. Since the only non-trivial strongly connected component in $\mathcal{A}$ is $\{q_a, q_b, q_c\}$, these states are the only candidates for

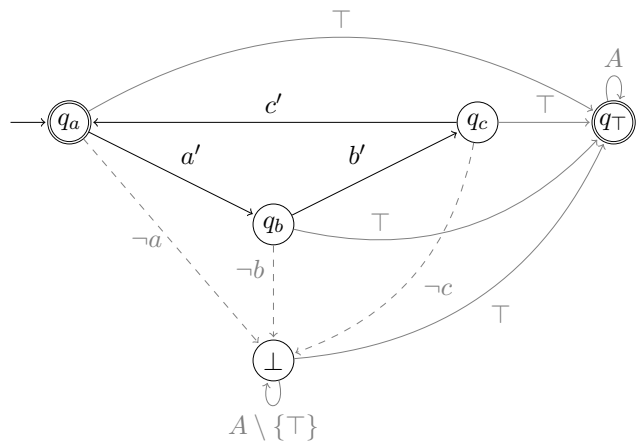


Figure 1: The minimal DFA $\mathcal{A}$ of K

p, q. Since p, q are distinct, it means $|u|$ is not a multiple of 3, and u induces a 3-cycle, either $q_a \xrightarrow{u} q_b \xrightarrow{u} q_c \xrightarrow{u} q_a$ if $|u| \equiv 1 \mod 3$ or in the reverse order if $|u| \equiv 2 \mod 3$. Thus, the first letter of u can be read from all states from $\{q_a, q_b, q_c\}$, while staying in this component. Such a letter does not exist, so we reach a contradiction. The DFA $\mathcal{A}$ is counter-free, so K is FO-definable [DG08]. □

4.3. **Syntactic monoid for the language K.** It is instructive to see what the syntactic monoid of K looks like, in particular to get a first intuition on how an FO formula can be defined for K.

We depict this monoid M in Figure 2, using the eggbox representation based on Green's relations: boxes are $\mathcal{J}$-classes, lines are $\mathcal{R}$-classes, columns are $\mathcal{L}$-classes, and cells are $\mathcal{H}$-classes. See [Col11] for an introduction to Green's relations and eggbox representation.

The syntactic morphism $h : A^* \to M$ is easily inferred, as elements of the monoid in $h(A)$ are directly named after the letter mapping to them. The accepting part of M is $F = \{1, \binom{a}{b}\binom{b}{c}\binom{c}{a}, \binom{c}{a}\binom{a}{b}\binom{b}{c}, abc, \top\}$.

To show that K is FO-definable, it suffices to verify that M is aperiodic, which is directly visible on Figure 2, as all $\mathcal{H}$-classes are singletons (see [Col11]).

4.4. **Defining an FO formula for the language K.** We now give some intuition on how an FO-formula can recognize K.

Recall that $K = (a^\uparrow b^\uparrow c^\uparrow)^* + A^*\top A^*$. We describe here the behaviour of a formula witnessing that K is FO-definable.

The $A^*\top A^*$ part of K is just to rule out words containing $\top$ by accepting them, which can be done by a formula $\exists x.\top(x)$. So we just need to design a formula φ for $K' = (a^\uparrow b^\uparrow c^\uparrow)^* \setminus (A^*\top A^*)$, assuming the letter $\top$ does not appear, the final formula will then be $\varphi \vee \exists x.\top(x)$.

We will call *forbidden pattern* any word that is not an infix of a word in K'. Let us call *anchor* a position x such that either x is labelled by a singleton, or x is labelled by $\binom{a}{b}$ (resp. $\binom{b}{c}, \binom{c}{a}$) with $x+1$ labelled by a letter different from $\binom{b}{c}$ (resp. $\binom{c}{a}, \binom{a}{b}$). The idea is that if

$\boxed{1}$

$\binom{a}{b}$	$\binom{a}{b}\binom{b}{c}$	$\binom{a}{b}\binom{b}{c}\binom{c}{a}$
$\binom{b}{c}\binom{c}{a}\binom{a}{b}$	$\binom{b}{c}$	$\binom{b}{c}\binom{c}{a}$
$\binom{c}{a}\binom{a}{b}$	$\binom{c}{a}\binom{a}{b}\binom{b}{c}$	$\binom{c}{a}$

a	ab	abc
bca	b	bc
ca	cab	c

$\boxed{\emptyset}$

$\boxed{\top}$

Figure 2: The syntactic monoid M of K

x is an anchor position of $u \in K'$, then there is only one possibility for the value of $x \mod 3$. If the first position is labelled by a letter from $a^{\uparrow}$, we will consider that it is an anchor labelled a, otherwise we will reject the input word. Similarly, the last position is either a c anchor or causes immediate rejection of the word. If x, y are successive anchor positions (i.e. with no other anchor positions between them), the word $u[x+1..y-1]$ is necessarily an infix of $(\binom{a}{b}\binom{b}{c}\binom{c}{a})^*$. We say that an anchor x *goes right-up* (resp. *right-down*) if we can replace the letter $\binom{\alpha}{\beta}$ by α (resp. β) at position $x+1$ without having a forbidden pattern in the immediate neighbourhood of x. Notice that x can not go both right-up and right-down. We define in the same way the left-up and left-down property by replacing $x+1$ with $x-1$. For instance consider $u = \binom{a}{b}\binom{b}{c}\binom{c}{a}\binom{a}{b}\binom{b}{c}c\binom{a}{b}\binom{b}{c}\binom{c}{a}\binom{a}{b}\binom{b}{c}\binom{b}{c}\binom{c}{a}\binom{a}{b}\binom{b}{c}$, then apart from the first and last position there are two anchors: $x = 5$ labelled c and $y = 10$ labelled $\binom{b}{c}$, because it is followed by another $\binom{b}{c}$.

$$\binom{a}{b}\binom{b}{c}\binom{c}{a}\binom{a}{b}\overset{x}{\binom{b}{c}c}\binom{a}{b}\binom{b}{c}\binom{c}{a}\binom{a}{b}\overset{y}{\binom{b}{c}}\binom{b}{c}\binom{c}{a}\binom{a}{b}\binom{b}{c}$$

Figure 3: A visualization of anchors

The anchor x goes left-up and right-up, while the anchor y goes left-up and right-down. If $d \in \{\text{up}, \text{down}\}$ is a direction, we say that two successive anchors $x < y$ *agree on* d if x goes right-d and y goes left-d. We say that x and y agree if they agree on some d.

Now, the formula φ will express the following properties:

- for all $x, x+1$ consecutive anchors, the letters at positions $x, x+1, x+2$ do not form a forbidden pattern (omit $x+2$ if $x+1$ is the last position).

- all non-consecutive successive anchors agree.

For instance the formula will accept the word u above, as the anchors $0, x$ agree on up, x, y agree on up, and $y, last$ agree on down.

It is routine to verify that these properties can be expressed in FO, and that they indeed characterize the language K'.

4.5. **Undefinability of K in FO^+.** To prove that K is the wanted counter-example, it remains to show:

Lemma 4.4. *K is not FO^+-definable.*

Proof. We establish this using Corollary 3.8. Let $n \in \mathbb{N}$, and $N = 2^n$. We define $u = (abc)^N$ and $v = [\binom{a}{b}\binom{b}{c}\binom{c}{a}]^{N-1}\binom{a}{b}\binom{b}{c}$. Notice that $u \in K$, and $v \notin K$ because $|v| \equiv 2 \bmod 3$, and v does not contain $\top$. By Corollary 3.8, it suffices to prove that $u \preceq_n v$ to conclude. We give a strategy for Duplicator in $\mathrm{EF}_n^+(u, v)$. The strategy is an adaptation from the classical strategy showing that $(aa)^*$ is not FO-definable [Lib04]. To simplify the description of the strategy, let us consider that prior to the game, tokens *first*, *last* are placed on the first and last positions on u, and *first*$'$, *last*$'$ on the first and last position of v. The strategy of Duplicator during the game is then as follows: every time Spoiler places a token in one of the words, Duplicator answers in the other by replicating the closest distance (and direction) to an existing token. This strategy is illustrated in Figure 4, where move i of Spoiler (resp. Duplicator) is represented by $\boxed{i}$ (resp. i).

$$\begin{array}{ccccccccccccccccccccccc}
 & & & & & & & & & \boxed{1} & & & 3 & & & & & & 2 & & & & \\
a & b & c & a & b & c & a & b & c & a & b & c & a & b & c & a & b & c & a & b & c & a & b & c \\
\binom{a}{b} & \binom{b}{c} & \binom{c}{a} & \binom{a}{b} & \binom{b}{c} & \binom{c}{a} & \binom{a}{b} & \binom{b}{c} & \binom{c}{a} & \binom{a}{b} & \binom{b}{c} & \binom{c}{a} & \binom{a}{b} & \binom{b}{c} & \binom{c}{a} & \binom{a}{b} & \binom{b}{c} & \binom{c}{a} & \binom{a}{b} & \binom{b}{c} & \binom{c}{a} & \binom{a}{b} & \binom{b}{c} \\
 & & & & & & & & & 1 & & & \boxed{3} & & & & & & \boxed{2} & & & & &
\end{array}$$

Figure 4: An example of Duplicator's strategy for $n = 3$.

Intuitively, the strategy of Duplicator is to match u with the top row of v if Spoiler plays close to the beginning of the words, and with the bottom row of v if Spoiler plays close to the end.

We have to show that this strategy of Duplicator allows him to play n rounds without losing the game. This proof is similar to the classical one for $(aa)^*$, see e.g. [Lib04], and actually the strategy is exactly the same if we forget the letter labels. The main intuition is that the length of the non-matching intervals between u and v is at worst divided by 2 at each round, and it starts with a length of 2^n, so Duplicator can survive n rounds.

Let us show that this strategy is indeed winning for Duplicator in $\mathrm{EF}_n^+(u, v)$.

We will generally write p, p' for related tokens, p being the position in u and p' the position in v.

The proof works by showing that the following invariant holds: after i rounds where Duplicator did not lose, if tokens in positions $p < q$ in u are related to tokens $p' < q'$ in v, and $u[p..q] \not\preceq_A v[p'..q']$, let us note $d = q - p, d' = q' - p'$; then $d = d' + 1$ and $d \geq 2^{n-i}$. In other words, if we call *wrong interval* a factor $u[p..q]$ or $v[p'..q']$ such that $u[p..q] \not\preceq_A v[p'..q']$, the invariant states that after i rounds, the length of the smallest wrong interval in u is at least 2^{n-i}, and corresponding wrong intervals differ by 1, the one in u being longer. Before

the first round, this invariant is true, as the only tokens are at the endpoints of u and v, and we have $|u| = |v| + 1$ and $|u| \geq 2^n$. Now, assume the invariant true at round i, and consider round $i+1$. When Spoiler plays a token in one of the words, two cases can happen. If it is played between previous tokens p, q (resp. p', q') such that $u[p..q] \leq_A v[p'..q']$, then Duplicator will simply answer the corresponding position in the other word, and the smallest wrong interval is not affected. If on the contrary, the new token is played in a minimal wrong interval, say $u[p, q]$ on position r, then Duplicator will answer by preserving the closest distance between $r - p$ and $q - r$. For instance if $r - p < q - r$, Duplicator will answer $r' = p' + (r - p)$. We can notice that by definition of the words u and v, and since $u[p] \leq v[p']$ by the rules of the game, we have $u[p..r] \leq_A v[p'..r']$, and in particular $u[r] \leq_A v[r']$, so the move of Duplicator is legal. Moreover, since $q - r > r - p$, we have $q - r \geq \frac{q-p}{2}$, so using the induction hypothesis, $q - r \leq 2^{n-(i+1)}$. Moreover, since we had $(q - p) = (q' - p') + 1$, we now have $(q - r) = (q - p) - (r - p) = (q' - p') + 1 - (r' - p') = (q' - r') + 1$, so the invariant is preserved. The case where $r - p \geq q - r$ is symmetrical. If on the other hand Spoiler plays in v a position r' in a wrong interval $v[p'..q']$, then $\min(r' - p', q' - r')$ will be strictly smaller than $2^{n-(i+1)}$, and will be replicated by the answer r of Duplicator in $u[p..q]$. This means that the new smallest wrong interval created in u will have length at least $2^{n-(i+1)}$, thereby guaranteeing that the invariant is also preserved in this case. □

5. Lyndon's Theorem

In this section we will see what the existence of this counter-example language K means for Lyndon's theorem on other structures. We start by showing in Section 5.1 that it can be adapted to show the failure of Lyndon's theorem on finite structures. We then show in Section 5.3 that the counter-example can also be encoded in finite directed graphs, and finally on undirected graphs.

5.1. General structures.

We will consider here first-order logic on arbitrary signatures and unconstrained structures. All definitions of Section 3 can be naturally extended to this general setting, and all results from Section 3.2 extend straightforwardly. We will later extend the EF^+-game result as well.

Our goal is to see how Theorem 4.1 can be lifted to this general framework.

Definition 5.1. A formula φ is *monotone* in a predicate P if whenever a structure S is a model of φ, any structure S' obtained from S by adding tuples to P is also a model of φ.

Example 5.2. On graphs, where the only predicate is the edge predicate, the formula asking for the existence of a triangle is monotone, but the formula stating that the graph is not a clique is not monotone.

Definition 5.3. A formula φ is *positive* in P if it never uses P under a negation.

Let us recall the statement of Lyndon's Theorem, which holds on general (possibly infinite) structures:

Theorem 5.4 [Lyn59, Cor 2.1]**.** *If ψ is an FO formula monotone in predicates $P_1, \ldots, P_n$, then it is equivalent to a formula positive in predicates $P_1, \ldots, P_n$.*

We will now see explicitly how the language K from Section 4 can be used to show that Lyndon's Theorem fails on finite structures.

The failure of this theorem on finite structures was first shown in [AG87] with a very difficult proof, then reproved in [Sto95] with a simpler one, using the Ehrenfeucht-Fraïssé technique. Still, the proof from [Sto95] is quite involved compared to the one we present here.

Since Lyndon's theorem can be found in the literature under different formulations, and since it is not clear at first sight that they are equivalent, we make it clear here how the construction of this paper applies to all of them. This also serves the purpose of making explicit the exact signature needed in each formulation to show the failure on finite structures with our method.

We will describe a signature by its sequence of arities, and add a symbol $\uparrow$ to specify monotone predicates. For instance $(2, 1\uparrow)$ describes a signature consisting of a binary predicate and a monotone unary predicate. Notice that the order is not important here, we are only interested in the multiset of pairs arity/monotonicity.

(i) **Arbitrarily many monotone predicates**

This is the most general formulation of Lyndon's theorem, as made explicit in Theorem 5.4. Let us show that our language K shows its failure on finite structures.

We will use here the fact that if $P = \{a, b, c\}$ is a set of monadic predicates, then a finite model over the signature $(\leq, a, b, c)$ where the order $\leq$ is total is simply a finite word on the powerset alphabet $A = \mathcal{P}(P)$. Therefore, in order to view our words as general finite structures, it suffices to axiomatize the fact that $\leq$ a total order. This can be done with a formula $\psi_{tot} = (\forall x, y.\ x \leq y \vee y \leq x) \wedge (\forall x, y, z.\ x \leq y \wedge y \leq z \Rightarrow x \leq z) \wedge (\forall x, y.\ x \leq y \wedge y \leq x \Rightarrow x = y) \wedge (\forall x.\ x \leq x)$. Notice that ψ_{tot} is not monotone in the predicate $\leq$.

Let φ be the FO-formula defining K, obtained in Lemma 4.3, and let $\psi = \varphi \wedge \psi_{tot}$. Then, ψ is monotone in predicates a, b, c, and finite structures on signature $(\leq, a, b, c)$ satisfying ψ are exactly words of K. However, as we proved in Theorem 4.1, no first-order formula that is positive in predicates a, b, c can define the same class of structures, since the same formula interpreted on words would be an FO^+-formula for K.

This gives a counter-example on a signature $(2, 1\uparrow, 1\uparrow, 1\uparrow)$.

(ii) **Single monotone predicate**

Other formulations of Lyndon's Theorem use a single monotone predicate, as in [AG87] and [Sto95]. More precisely, [AG87] uses a signature $(1, 2, 3, 3, 3, 3, 3, 4, 1\uparrow)$, and [Sto95] a signature $(2, 2\uparrow)$.

We can encode the language K in this framework, by using one binary predicate A to represent all letter predicates. Let K_3 be K restricted to words of length at least 3. By Theorem 4.1 it is clear that K_3 is FO-definable but not FO^+-definable.

Let ψ_3 be an FO-formula stating that there are at least 3 elements $0, 1, 2$, and that for all $y \notin \{0, 1, 2\}$ and for all x, $A(x, y)$ holds. We build the FO formula φ' from the FO formula φ recognizing the language K_3 by replacing every occurrence of $a(x)$ (resp. $b(x), c(x)$) by $A(x, 0)$ (resp. $A(x, 1), A(x, 2)$).
Finally, we define the FO formula $\psi' = \psi_{tot} \wedge \psi_3 \wedge \varphi'$. Finite structures on signature $(\leq, A)$ accepted by ψ' are exactly those which encode words of K_3. No formula positive in A can recognize this class of structures, otherwise we could obtain from it an FO^+-formula for K_3, by replacing every occurrence of $A(x, y)$ by $(a(x) \wedge y = 0) \vee (b(x) \wedge y = 1) \vee (c(x) \wedge y = 2) \vee y \geq 3$.

We thus give a counter-example on a signature $(2, 2\uparrow)$, as was done in [Sto95].

(iii) **Closure under surjective homomorphisms**

Lyndon's theorem is also often stated in the following way: if an FO formula defines a class of structures closed under surjective homomorphisms, then it is equivalent to a positive formula. This formulation is equivalent to saying that the formula is monotone in all predicates. This case has been treated in [Ros95], using a slight modification of the construction from [Sto95], and building a counter-example on signature $(0, 0, 1\uparrow, 1\uparrow, 2\uparrow, 2\uparrow, 2\uparrow)$. Notice that arities 0 correspond to constants, which are always trivially monotone.

We can deal with this framework as well, by incorporating a predicate $\not\leq$ to the signature. Let $\psi_{\not\leq}$ be the formula obtained from ψ defined in case (i) by pushing negations to the leaves and replacing all subformulas of the shape $\neg(x \leq y)$ with $x \not\leq y$. Let

$$\begin{aligned} \psi_- &= \forall x, y.\ (x \leq y \vee x \not\leq y) \\ \psi_+ &= \exists x, y.\ (x \leq y \wedge x \not\leq y) \\ \psi'' &= (\psi_- \wedge \psi_{\not\leq}) \vee \psi_+. \end{aligned}$$

Finite structures on the signature $(\leq, \not\leq, a, b, c)$ can be classified into three categories:

(1) if there are x, y such that $x \leq y \wedge x \not\leq y$, then the structure satisfies ψ_+ hence ψ''
(2) otherwise, if there are x, y such that $\neg(x \leq y \vee x \not\leq y)$, then the structure does not satisfy ψ_- so it does not satisfy ψ'' either.
(3) otherwise, $\not\leq$ is the complement of $\leq$, and the structure satisfies ψ'' if and only if it satisfies $\psi_{\not\leq}$.

Therefore, in $\psi_{\not\leq}$ we can use $\leq$ and $\not\leq$ freely, assuming that $\not\leq$ is actually the complement of $\leq$. In particular the ψ_{tot} subformula of $\psi_{\not\leq}$ axiomatizes the fact that $\leq$ is a total order, provided that $\not\leq$ is its complement. So the structures of item (3) are exactly the words of K, with an additional predicate $\not\leq$ which is the complement of $\leq$. Items (1),(2) guarantee that the class of finite structures accepted by ψ'' is monotone with respect to $\leq$ and $\not\leq$ as well. As before, it is impossible to have a formula positive in all predicates accepting the same class of finite structures as ψ'', since replacing $x \not\leq y$ with $y < x$ in this formula would directly yield an FO^+-formula for K.

We thus give a counter-example on signature $(2\uparrow, 2\uparrow, 1\uparrow, 1\uparrow, 1\uparrow)$, which can also be adapted to a signature $(2\uparrow, 2\uparrow, 2\uparrow)$ by applying the trick of case (ii). Both results constitute an improvement upon the signature from [Ros95].

5.2. **EF^+-games on arbitrary signature.** We consider here FO^+ on an arbitrary signature consisting of l monotone predicates $P_1, \ldots, P_l$ of arity $r_1, \ldots, r_l$ respectively, and l' non-monotone predicates $R_1, \ldots, R_{l'}$ of arity $r'_1, \ldots, r'_{l'}$ respectively.

Thus the syntax of FO^+ in this setting is:

$$\varphi, \psi := P_i(x) \mid R_i(x) \mid \neg R_i(x) \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \exists x.\varphi \mid \forall x.\varphi$$

Notice that we allow the negation of predicates from R_i.

We can additionally assume that these formulas will only be evaluated on structures verifying certain axioms, for instance on structures where a predicate $\leq$ evaluates to a linear order, or where the predicate $=$ corresponds to equality. These will be called σ-structures in the following, where the signature σ can be enriched by such axioms.

The EF^+-game on two σ-structures u, v is defined as before, with the following generalization: at a given stage (u, α, v, β) of the game, where α (resp. β) is a valuation in u (resp. v) for already played tokens, Duplicator must ensure:

- For any monotone predicate P_i and tuple $\vec{x}$ of r_i played tokens, we must have $u, \alpha \models P_i(\vec{x}) \Rightarrow v, \beta \models P_i(\vec{x})$.
- For any non-monotone predicate R_i and tuple $\vec{x}$ of r'_i played tokens, we must have $u, \alpha \models P_i(\vec{x}) \Leftrightarrow v, \beta \models P_i(\vec{x})$.

As before, we note $u \preceq_n v$ if Duplicator has a strategy to win the n-round EF^+-game between s and t. We want to show the following theorem, generalizing Theorem 3.7, and formulated in [Sto95]:

Theorem 5.5 [Sto95, Thm 2.4]**.** *In a general setting, for any σ-structures u, v, we have $u \preceq_n v$ if and only if for all formulas φ of FO^+ with $\mathrm{qr}(\varphi) \leq n$, we have $(s \models \varphi) \Rightarrow (t \models \varphi)$.*

The proof is an adaptation of the classical proof for correctness of EF-games, see e.g. [Lib04].

Since FO^+ is a fragment of FO, we can directly use the following Lemma:

Lemma 5.6 [Lib04, Lem 3.13]**.** *Let $n, k \in \mathbb{N}$. Up to logical equivalence, there are finitely many formulas of quantifier rank at most n using k free variables.*

We will now show a strengthening of Theorem 5.5, where free variables are incorporated in the statement:

Theorem 5.7. *Let $n, k \in \mathbb{N}$, u, v σ-structures, $\alpha : [1, k] \to dom(u)$ and $\beta : [1, k] \to dom(v)$ be valuations for k variables $x_1, \ldots, x_k$ in u, v respectively. Then Duplicator wins $\mathrm{EF}_n^+(u, \alpha, v, \beta)$ if and only if for any FO^+ formula φ with $\mathrm{qr}(\varphi) \leq n$ using k free variables $x_1 \ldots x_k$, we have $u, \alpha \models \varphi \;\Rightarrow\; v, \beta \models \varphi$.*

Proof. We prove this by induction on n.
Base case $n = 0$:

Notice that quantifier-free formulas of FO^+ are just positive boolean combinations of atomic formulas of the form $P_i(\vec{x})$, $R_i(\vec{x})$ or $\neg R_i(\vec{x})$. We will note $Q(\vec{x})$ for such an arbitrary atomic formula. Let φ be such a formula with k free variables accepting u, α but rejecting v, β. This happens if and only if there is an atomic formula $Q(\vec{x})$ such that $u, \alpha \models Q(\vec{x})$ and $v, \beta \not\models Q(\vec{x})$. This is equivalent to saying that (u, α, v, β) is not a valid k-position, i.e. Spoiler wins the 0-round game $\mathrm{EF}_0^+(u, \alpha, v, \beta)$.

Induction case: Assume there is an FO^+ formula φ with $\mathrm{qr}(\varphi) \leq n$, accepting u, α but not v, β. The formula φ is a positive combination of atomic formulas, formulas of the form $\exists x.\psi$, and formulas of the form $\forall x.\psi$. Therefore, one of these formulas accepts u, α but not v, β. If it is an atomic formula, then Spoiler immediately wins $\mathrm{EF}_n^+(u, \alpha, v, \beta)$ as in the base case.

If it is a formula of the form $\exists x.\psi$, then Spoiler can use the following strategy: pick a position p witnessing that the formula is true for u, α, and play the position p in u. Duplicator will answer a position p' in v, and the game will move to (u, α', v, β'), where $\alpha' = \alpha[x \mapsto p]$ and $\beta' = \beta[x \mapsto p']$. Since the formula ψ has quantifier rank at most $n - 1$, and accepts u, α' but not v, β', by induction hypothesis Spoiler can win in the remaining $n - 1$ rounds of the game.

Now if it is a formula of the form $\forall x.\psi$, then Spoiler can do the following: pick a position p' witnessing that the formula is false for v, β, and play the position p' in v. Duplicator will answer a position p in u, and the game will move to (u, α', v, β'), where $\alpha' = \alpha[x \mapsto p]$ and $\beta' = \beta[x \mapsto p']$. Since the formula ψ has quantifier rank at most $n - 1$, and accepts u, α' but not v, β', by induction hypothesis Spoiler can win in the remaining $n - 1$ rounds of the game.

Let us now show the converse implication. We assume any formula of quantifier rank at most n accepting u, α must accept v, β, and we give a strategy for Duplicator in $\mathrm{EF}_n^+(u, \alpha, v, \beta)$.

Suppose Spoiler places token x at position p in u. Let $\alpha' = \alpha[x \mapsto p]$. By Lemma 5.6, up to logical equivalence, there is only a finite set F of FO^+ formulas of rank at most $n-1$ with $k+1$ free variables accepting u, α'. Let $\psi = \bigwedge_{\varphi \in F} \varphi$. Then u, α satisfies the formula $\exists x.\psi$ of rank n (as witnessed by p), so by assumption we also have $v, \beta \models \exists x.\psi$. This means there is a position p' of v such that $v, \beta' \models \psi$, where $\beta' = \beta[x \mapsto p']$. Duplicator can answer position p' in v, and by induction hypothesis he will win the remaining of the game, since every formula of F accepts v, β'.

Suppose now that Spoiler places token x at position p' in v. Let $\beta' = \beta[x \mapsto p']$. Let F be the finite set of formulas (up to equivalence) of quantifier rank at most $n-1$ and with $k+1$ free variables, that reject v, β'. Let $\psi = \bigvee_{\varphi \in F} \varphi$, and $\psi' = \forall x.\psi$. By construction, $x = p'$ witnesses that ψ' does not accept v, β. Our assumption implies that it does not accept u, α either. So there is $p \in dom(u)$ such that $u, \alpha' \not\models \forall x.\psi$, where $\alpha' = \alpha[x \mapsto p]$. Duplicator can answer position p in u. If a formula φ of rank at most $n-1$ is true in u, α', then by construction it cannot appear in F, therefore it is also true in v, β'. By induction hypothesis, Duplicator wins the remaining $(n-1)$-round game starting from (u, α', v, β'). □

This achieves the proof of Theorem 5.5, and its instantiation Theorem 3.7 on finite words. The proof of Corollary 3.8 is exactly identical in this general setting, so EF^+ game can be used to prove that some properties of general structures are not expressible in FO^+:

Corollary 5.8. *A class of σ-structures C is not FO^+-definable if and only if for all $n \in \mathbb{N}$, there exists $(u, v) \in C \times \overline{C}$ such that $u \preceq_n v$.*

5.3. **Finite directed graphs.** Our goal is now to show that Lyndon's theorem fails on finite directed graphs, i.e. on finite structures where the signature consists in one (monotone) binary predicate, in addition to (non-monotone) equality. To our knowledge this is a new result.

The positive FO formulas on graphs, that we will call again FO^+, is defined via the following syntax:

$$\varphi, \psi := E(x, y) \mid x = y \mid x \neq y \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \exists x.\varphi \mid \forall x.\varphi$$

while general FO formulas can additionally use predicates of the form $\neg E(x, y)$. A class $\mathcal{C}$ of graphs is *monotone* if whenever $G \in \mathcal{C}$, and G' is obtained from G by adding edges, then $G' \in \mathcal{C}$. It is straightforward to adapt the proof of Lemma 3.5 to show that FO^+ can only define monotone classes of graphs.

Notice that equality/inequality predicates were not needed in the case of words since this was expressible with the order predicates $\leq$ and $<$.

The goal of this section is to prove the following result:

Theorem 5.9 (Failure of Lyndon's Theorem on finite directed graphs)**.** *There exists an FO-definable monotone class of directed graphs, which is not FO^+-definable.*

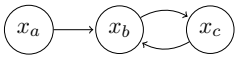
Let us start by giving an informal proof sketch. Our goal will simply be to encode the language K from Section 4 as a set of graphs. Thus, the graphs of interest will have a very specific shape, allowing to encode words on alphabet $A = \mathcal{P}(\{a, b, c\})$. In order to

ensure monotonicity, instead of forbidding all patterns that break the encoding, we will instead accept any graph having "too many edges". This is the same idea as in the "Closure under surjective homomorphisms" paragraph of Section 5.1. Thus we will have two kinds of constraints:

- A formula ψ_- asking for some edges to be present, i.e. rejecting graphs that cannot encode a word because of a lack of edges.
- A formula ψ_+ that will accept any graph falling outside of the required shape of an encoding, because of an excess of edges.

Both ψ_- and ψ_+ will be FO^+ formulas, thus describing monotone classes of graphs. The graphs encoding words of A^* will be the models of ψ_- that are not models of ψ_+. Let us call $\mathcal{G}_w$ the set such graphs, encoding words of A^*. We call $\mathcal{G}_K$ the subset of $\mathcal{G}_w$ consisting of graphs encoding words of K. Our monotone language of graphs witnessing failure of Lyndon's theorem will be $\mathcal{G}_K \cup [\![\psi_+]\!]$. Let $(\varphi_K)^G$ be an FO formula accepting $\mathcal{G}_K$ among graphs from $\mathcal{G}_w$, the behaviour of $(\varphi_K)^G$ being irrelevant outside of $\mathcal{G}_w$. This formula $(\varphi_K)^G$ will be obtained from the FO formula φ_K for the language K, by interpreting predicates $a(x)$, $b(x)$, $c(x)$, and $x \leq y$ in our encoding. For technical reasons, our encoding will also use some distinguished vertices $\vec{x}$, shared by all formulas, and existentially quantified. Thus our final formula will be of the form $\phi := \exists \vec{x}.\psi_- \wedge ((\varphi_K)^G \vee \psi_+)$. It accepts a monotone language of graphs: $\mathcal{G}_K \cup [\![\psi_+]\!]$. In order to show that there is no positive formula equivalent to ϕ, we will replicate the EF^+ game of Lemma 4.4 using graphs from $\mathcal{G}_w$.

Let us now move to the detailed construction. Let $E(x, y)$ be the edge predicate, that we will often note $x \to y$ for simplicity. Similarly, we will note $x \to y \to z$ as a shortcut for $E(x, y) \wedge E(y, z)$. This arrow will not be at risk of being confused with an implication symbol, since we will avoid the use of implication to obtain negation-free formulas. In the following, we assume three vertices x_a, x_b, x_c are pointed in the graph. We will call them *sources*, represented by circles in figures. We will call "squares" the vertices other than x_a, x_b, x_c. We will impose that the subgraph induced by the sources is a particular one, the only purpose of this is to be able to uniquely identify these three vertices. We define $\mathcal{G}_w$ to be the set of graphs satisfying the following properties:

(sources) x_a, x_b, x_c are distinct, and they induce the following subgraph: 

(in-edge) x_a is the only vertex with no in-edge
(cycle) There is no cycle of length at most 3 other than the 2-cycle on x_b, x_c.
(order) Any two squares are related by an edge.
(direction) There is no edge from a square to a source.

The rule (direction) is actually optional for the correctness of the construction, but it simplifies the exposition. The next lemma justifies the choice of name for the rule (order):

Lemma 5.10. *If G is a graph in $\mathcal{G}_w$, the edge relation defines a strict total order on squares.*

Proof. Since 3-cycles are forbidden, and any two squares are related, the edge relation is transitive on squares: for any $x \to y \to z$, there is an edge $x \to z$. Since self-loops and 2-cycles are forbidden, the relation is irreflexive and antisymmetric. Therefore, it defines a strict total order on squares. □

Figure 5 shows the shape of such a graph with four squares. Notice that the edges from sources to squares can be arbitrary, except that there must be an edge from a source to the first square in the order, because of rule (in-edge).

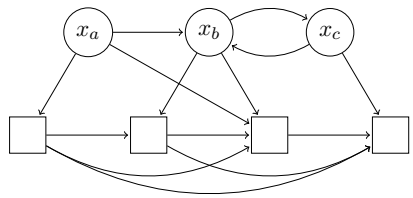


Figure 5: A graph of $\mathcal{G}_w$

Let us give at this stage more details about sources: their role will be to encode the unary predicates a, b, c on words, and the constraints of $\mathcal{G}_w$ allow to identify them without ambiguity in an unlabeled graph: x_a is the only vertex with no in-edge, and x_b, x_c form the only 2-cycle, x_b being the vertex connected to x_a.

We now need to express all these constraints via FO$^+$ formulas ψ_- and ψ_+. Recall that formulas in ψ_+ are meant to express when a graph is not in $\mathcal{G}_w$ because of extra edges. Let $E_\circ = \{(x_a, x_b), (x_b, x_c), (x_c, x_b)\}$ the set of edges in the induced graph of rule (source), and $\overline{E_\circ} = \{x_a, x_b, x_c\}^2 \setminus E_\circ$ its complement. We will also use as auxiliary formulas:

- $\bigcirc(x) := (x = x_a) \vee (x = x_b) \vee (x = x_c)$, stating that x is a source,
- $\square(x) := (x \neq x_a) \wedge (x \neq x_b) \wedge (x \neq x_c)$ stating that x is a square,
- $\square_a(x) := (x \neq x_b) \wedge (x \neq x_c)$ for squares or x_a.

Constraint	$\boldsymbol{\psi_-}$	$\boldsymbol{\psi_+}$
(sources)	$\bigwedge_{(x,y)\in E_\circ}(x \to y \wedge x \neq y)$	$\bigvee_{(x,y)\in\overline{E_\circ}} x \to y$
(in-edge)	$\forall y.(y = x_a \vee \exists x.x \to y)$	$\exists x.x \to x_a$
(cycle)		$\exists x.x \to x$ $\vee\ \exists x, y.(\square_a(x) \vee \square_a(y)) \wedge (x \to y \to x)$ $\vee\ \exists x, y, z.(x \to y \to z \to x)$
(order)	$\forall x, y.\ \bigcirc(x) \vee \bigcirc(y) \vee (x = y)$ $\vee (x \to y) \vee (y \to x)$	
(direction)		$\exists x, y.\square(x) \wedge \bigcirc(y) \wedge x \to y$

We finally take for the ψ_- (resp. ψ_+) the conjunction (resp. disjunction) of the formulas in its column. We obtain that by definition, a graph with marked vertices x_a, x_b, x_c is in $\mathcal{G}_w$ if and only if it satisfies ψ_- but not ψ_+.

It remains to describe how words on alphabet $A = \mathcal{P}(\{a, b, c\})$ can be encoded into these graphs. Let $G \in \mathcal{G}_w$, we will associate to it a word u using the following rules:

- The positions $dom(u)$ of u correspond to the square vertices of G.
- The order $<$ on $dom(u)$ corresponds to edges between square vertices.
- If $x \in dom(u)$, we will say that $a(x)$ (resp. $b(x), c(x)$) is true if there is in G an edge $x_a \to x$ (resp. with x_b, x_c).

With this, we can see that the graph of Figure 5 encodes the word $ab\binom{a}{b}c$.

Notice that this encoding is actually a bijection between graphs of $\mathcal{G}_w$ and words in A^* that do not start with letter $\emptyset$.

We note $\mathcal{G}_K$ the graphs of $\mathcal{G}_w$ that encode a word $u \in K$, where K is the language from Section 4. Recall that letter $\emptyset$ is never allowed in a word of K. Let φ_K be the FO-formula for K (from Lemma 4.3), on signature $(\leq, a, b, c)$. We want to build a formula $(\varphi_K)^G$,

recognizing the graphs of $\mathcal{G}_K$ among those of $\mathcal{G}_w$. The idea is to restrict quantification to square vertices, and replace atomic predicates by their graph interpretations. This is done by induction on formulas:

- $(x \leq y)^G := E(x, y) \vee x = y$
- $(x < y)^G := E(x, y)$
- $(\alpha(x))^G := E(x_\alpha, x)$, for $\alpha \in \{a, b, c\}$.
- $(\exists x.\varphi)^G := \exists x.\Box(x) \wedge \varphi^G$
- $(\forall x.\varphi)^G := \forall x. \bigcirc (x) \vee \varphi^G$
- $(\varphi \wedge \psi)^G := \varphi^G \wedge \psi^G$
- $(\varphi \vee \psi)^G := \varphi^G \vee \psi^G$
- $(\neg\varphi)^G := \neg(\varphi^G)$

Notice that since φ_K is not syntactically positive, the formula $(\varphi_K)^G$ will contain negations as well.

Finally, let us define

$$\phi := \exists x_a, x_b, x_c.(\psi_- \wedge ((\varphi_K)^G \vee \psi_+)).$$

Remark 5.11. Contrarily to the last construction of Section 5.1, how to parenthesize is important here, because of the existential quantification on sources x_a, x_b, x_c. Indeed, we must avoid being too permissive by allowing a choice of sources that would validate ψ_+ without validating ψ_-, thereby accepting some unwanted graphs because of a bad choice of sources. Actually, with the other choice of parentheses $(\psi_- \wedge (\varphi_K)^G) \vee \psi_+$, we would have ϕ accepting any graph containing an edge, because of rule (in-edge) in ψ_+.

Lemma 5.12. *The formula ϕ is monotone, and $[\![\phi]\!] \cap \mathcal{G}_w = \mathcal{G}_K$.*

Proof. We first show that $[\![\phi]\!] \cap \mathcal{G}_w = \mathcal{G}_K$. Remark that given a graph in $\mathcal{G}_w$, there is only one possible choice of x_a, x_b, x_c that validates ψ_-: they are the only vertices forming the subgraph given in rule (sources). Thus, after x_a, x_b, x_c have been fixed, the fact that the graphs of $\mathcal{G}_w$ accepted by ϕ are exactly those from $\mathcal{G}_K$ follows from this: the formula $(\varphi_K)^G$ interprets correctly a graph of $\mathcal{G}_w$ as a word of A^*, and accepts it if and only if this word is in K. This fact is simply the correctness of the $(\cdot)^G$ transformation, proven by straightforward induction.

We now move to the monotonicity property: let $G \in [\![\phi]\!]$, and G' be obtained from G by adding some edges.

First, if $G \in [\![\psi_+]\!]$, then so does G', because ψ_+ is positive. If on the contrary $G \notin [\![\psi_+]\!]$, this means $G \in \mathcal{G}_K$, so G encodes a word $u \in K$. Then two cases can occur for G':

- If $G' \in \mathcal{G}_w$, then it encodes a word $u' \geq_A u$, because adding edges to a graph of $\mathcal{G}_w$ while staying inside $\mathcal{G}_w$ translates into adding letter predicates in the corresponding word. Since K is monotone, we have $u' \in K$, so $G' \in \mathcal{G}_K$.
- If $G' \notin \mathcal{G}_w$, recall that $\mathcal{G}_w = [\![\psi_-]\!] \setminus [\![\psi_+]\!]$. Since $G \in [\![\psi_-]\!]$ and ψ_- is positive, we have $G' \in [\![\psi_-]\!]$. So we can conclude $G' \in [\![\psi_+]\!]$.

In all cases, we obtain $G' \in [\![\phi]\!]$, so ϕ is monotone. □

Lemma 5.13. *There is no* FO^+ *formula recognizing* $[\![\phi]\!]$.

Proof. We use EF^+-games on graphs of $\mathcal{G}_w$. By Corollary 5.8, we can use EF^+-games on signature $(E, =)$, with E monotone and $=$ non-monotone and correspond to equality, to show that there is no positive formula defining $[\![\phi]\!]$. We will replicate the game from Lemma 4.4,

using graphs instead of words. More precisely, let $n \in \mathbb{N}$ and $N = 2^n$. Let $u = (abc)^N$ and $v = [\binom{a}{b}\binom{b}{c}\binom{c}{a}]^{N-1}\binom{a}{b}\binom{b}{c}$, recall that $u \in K$ and $v \notin K$. We associate to u and v graphs G_u and G_v in $\mathcal{G}_w$ according to the above encoding. We have $G_u \in \mathcal{G}_K$ and $G_v \in \mathcal{G}_w \setminus \mathcal{G}_K$, so $G_u \in [\![\phi]\!]$ and $G_v \notin [\![\phi]\!]$. It now suffices to show that Duplicator can win $\mathrm{EF}_n^+(G_u, G_v)$. His strategy can actually mimic exactly the strategy σ_D for Duplicator in $\mathrm{EF}_n^+(u, v)$. It suffices to play as σ_D on squares, and if Spoiler plays a source in one of the graphs, Duplicator answers with the corresponding source in the other. It is straightforward to verify that this is a winning strategy for Duplicator in $\mathrm{EF}_n^+(G_u, G_v)$, as both order and letter predicates on words directly translate to edge predicates on graphs of $\mathcal{G}_w$. This shows that there is no FO^+ formula equivalent to ϕ. □

This concludes the proof of Theorem 5.9.

5.4. **Finite undirected graphs.** We describe in this section how to lift the counter-example from directed graphs to undirected ones.

As the proof scheme follows the same pattern as in the directed case, we will go in less details and just describe here the modifications to be made to lift the previous proof to undirected graphs.

We will again use distinguished "source vertices" that will encode letter predicates, except that now there will be more than one source per letter. As we can no longer use the orientation of edges to isolate these sources without ambiguity, we will instead use cycles, in the same spirit as the $x_b - x_c$ cycle in the directed case: the sources will form the only cycles of length at most 5. More precisely, the a-sources will form a 3-cycle, the b-sources a 4-cycle, and the c-sources a 5-cycle, for a total of 12 sources. Moreover, there are no other edges between sources than the ones forming these cycles. This means that once again we completely impose the graph induced by the 12 sources: it has to consist in three disjoint cycles of size $3, 4, 5$. We can therefore assume that we have a formula $\bigcirc(x)$ stating that x is a source, and formulas $\bigcirc_a(x), \bigcirc_b(x), \bigcirc_c(x)$ specifying the letter of this source. Since sources will again be explicitly quantified in a formula, it is still possible to state in FO^+ that a vertex x is not a source, via a formula $\neg \bigcirc (x)$ simply asserting that x is different from all sources.

As before, we will use vertices called "squares" to encode positions of the word. This time, squares will not be all non-source vertices, but will be defined as follows: a square is any non-source vertex that is connected to a source by an edge. This can be expressed by a formula $\square(x) = \neg \bigcirc (x) \wedge \exists y. \bigcirc (y) \wedge E(x, y)$. We still need to encode the total order on squares, but since edges are not oriented anymore, we will make use of oriented "meta-edges" as described by the following picture:

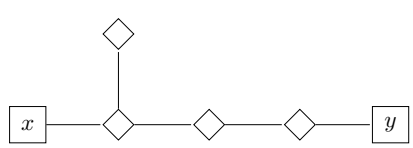


Figure 6: A meta-edge from square x to square y

The fact that there is a meta-edge from square x to square y can be described by a formula $M(x, y)$ asserting the existence of the above pattern. Notice that such a meta-edge

can create a cycle of length 6, if x and y are connected to the same source, but will not introduce a cycle of length at most 5.

We call "diamonds" the auxiliary vertices that are not sources, and that are connected to a square by a path of length 1 or 2. This can be defined by a positive formula $\Diamond(x)$.

For a graph to be in $\mathcal{G}_w$, we require the following additional constraints:

(partition) Any vertex is either a source, a square or a diamond, with no overlap.
(cycle) The only cycles of length at most 5 are those composed of sources.
(order) The meta-edge relation form a strict total order on squares.
(diamonds) A diamond can be part of at most one meta-edge.

As before, we can express all these constraints with positive formulas ψ_- and ψ_+. Let us explicit these formulas for rules (partition) and (order). We can use $\neg\Box(x) := \bigcirc(x) \vee \Diamond(x)$ to assert that x is not a square, thanks to rule (partition). This allows us to quantify on squares only, that we will abbreviate $\exists^\Box$ and $\forall^\Box$.

Constraint	$\boldsymbol{\psi_-}$	$\boldsymbol{\psi_+}$
(partition)	$\forall x. \bigcirc(x) \vee \Box(x) \vee \Diamond(x)$	$\exists x. \Box(x) \wedge \Diamond(x)$
(order)	$\forall^\Box x, y. M(x,y) \vee M(y,x)$	$\exists^\Box x. M(x,x)$ $\vee\ \exists^\Box x, y. M(x,y) \wedge M(y,x)$ $\vee\ \exists^\Box x, y, z. M(x,y) \wedge M(y,z) \wedge M(z,x)$

The formulas for (cycle) and (diamonds) pose no additional difficulty: the one for (cycle) is similar to the directed case, and the one for (diamonds) only has a ψ_+ component, stating the existence of two meta-edges sharing a diamond. The formula ψ_- (resp. ψ_+) will again be the conjunction (resp. disjunction) of its components coming from various rules.

A graph $G \in \mathcal{G}_w$ will be an encoding of a word $u \in (A \setminus \{\emptyset\})^*$, by interpreting the squares with the meta-edge order as $(dom(u), <)$, and a predicate $a(x)$ true if the square x is connected to an a-source (resp. with b, c).

For instance the following graph encodes the word $ab\binom{a}{b}c$, where meta-edges are represented by dashed arrows:

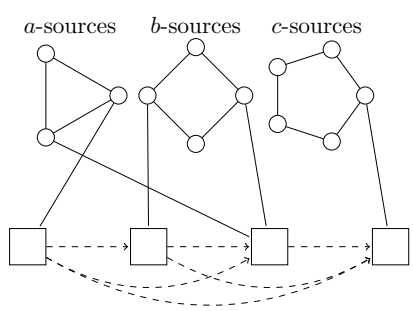


Figure 7: A graph of $\mathcal{G}_w$ encoding $ab\binom{a}{b}c$

Similarly to the directed case, we will transform the formula φ_K into a formula $(\varphi_K)^G$ using the following translation:

- $(x \leq y)^G := M(x,y) \vee x = y$
- $(x < y)^G := M(x,y)$

- $(\alpha(x))^G := \exists x_\alpha. \bigcirc_\alpha (x_\alpha) \wedge E(x_\alpha, x)$, for $\alpha \in \{a, b, c\}$.
- $(\exists x.\varphi)^G := \exists^\square x.\varphi^G$
- $(\forall x.\varphi)^G := \forall^\square x.\varphi^G$
- $(\varphi \wedge \psi)^G := \varphi^G \wedge \psi^G$
- $(\varphi \vee \psi)^G := \varphi^G \vee \psi^G$
- $(\neg\varphi)^G := \neg(\varphi^G)$

As before, we define $\phi = \exists\vec{x}.(\psi_- \wedge ((\varphi_K)^G \vee \psi_+))$, where $\vec{x}$ is the list of the 12 source vertices. The fact that ϕ is monotone and $[\![\phi]\!] \cap \mathcal{G}_w = \mathcal{G}_K$, the graphs encoding words in K, is proved the same way.

We again show that there is no positive formula for ϕ using the EF^+-game technique according to Corollary 5.8. Given words u and v as before, we choose canonical encodings G_u and G_v in $\mathcal{G}_w$, where all diamonds are used in some meta-edge. We choose G_u and G_v of the simplest possible form, for instance without parallel meta-edges. We will show that Duplicator can use his winning strategy from $\mathrm{EF}_n^+(u, v)$ to win in the (G_u, G_v) arena as well. As before, sources and squares pose no problem, and the strategy can be directly copied from words to graphs. However, there is a new subtlety to take care of: Spoiler can now play on diamond vertices in G_u or G_v. Since by rule (diamonds), any diamond is part of exactly one meta-edge from a square x to a square y. In order to answer this move in the graph game, Duplicator will look at what happens in the word game if Spoiler plays positions x and y. Duplicator's winning strategy gives answers x' and y' to these moves. He can now answer the corresponding diamond in the other graph, in the meta-edge between squares x' and y', at the same relative position as the diamond originally played by Spoiler. Since playing 2 moves on words can be necessary to answer one move on graphs, Duplicator will only be able to play $n/2$ rounds in the game on G_u, G_v. This is enough to show that for any n there are $G_u \in [\![\phi]\!]$ and $G_v \notin [\![\phi]\!]$ such that Duplicator wins $\mathrm{EF}_n^+(G_u, G_v)$, thereby proving that $[\![\phi]\!]$ is not definable in FO^+.

This concludes the proof scheme of the following theorem:

Theorem 5.14. *Lyndon's Theorem fails on finite undirected graphs: there is an FO-definable monotone class of undirected graphs that is not definable with a negation-free formula.*

6. Undecidability of FO^+-definability for regular languages

In this section, we are back to considering only finite words. Our goal will be to prove the following Theorem:

Theorem 6.1. *The following problem is undecidable: given L a regular language on a powerset alphabet, is L FO^+-definable?*

Notice that since FO-definability is decidable for regular languages, this theorem could be equivalently stated with an input language given by an FO formula.

We will a start with an informal proof sketch to convey the main ideas of the proof, before going to the technical details.

6.1. **Proof sketch.** The proof proceeds by reduction from the Turing Machine Mortality problem, known to be undecidable [Hoo66]. A deterministic Turing Machine (TM) is *mortal* if there is a uniform bound $n \in \mathbb{N}$ on the length of its runs, starting from any arbitrary configuration.

Given a machine M, we want to build a regular language L_M such that L_M is FO^+-definable if and only if M is mortal.

Configuration words

The intuitive idea is that L_M will mimic the language $(a^\uparrow b^\uparrow c^\uparrow)^*$ from Definition 4.2, but the letters will be replaced by words encoding configurations of M. We therefore design an ordered alphabet A and a language C of configurations words such that words from C encode configurations of M. These words will be of three possible types $1, 2, 3$, playing the role of the letters a, b, c of the language K. This partitions C into $C_1 \cup C_2 \cup C_3$. We guarantee that the transitions of M will always change the types in the following way: $1 \to 2$, $2 \to 3$ or $3 \to 1$.

Moreover, we design the order $\leq_A$ of the alphabet A so that given two words u_1, u_2 from C, there is a word v that is bigger (for the order $\leq_A$) than both u_1 and u_2 if and only if u_1 and u_2 are consecutive configurations of M. Such a word v will be written $\binom{u_1}{u_2}$ in this proof sketch.

Language L_M

Finally, the language L_M will be roughly defined as the upward-closure of $(C_1\#C_2\#C_3\#)^*$, where # is a separator symbol. The only requirement for configuration words appearing in a word of L_M is on their types. Apart from this, the configuration words can be arbitrary, they do not have to form a run of M.

We will then use the EF^+-game technique to show that L_M is FO^+-definable if and only if M is mortal.

If M not mortal

The easier direction is proving that if M is not mortal, then L_M is not FO^+-definable. Indeed, if M is not mortal, we can choose an arbitrarily long run $u = u_1\#u_2\#\ldots\#u_N$ of M, where $u_1 \in C_1$ and N is a multiple of 3 (implying $u_N \in C_3$). We build the word $v = \binom{u_1}{u_2}\#\binom{u_2}{u_3}\#\ldots\#\binom{u_{N-1}}{u_N}$, and we verify that $u \in L_M$ and $v \notin L_M$ (because of the number of # modulo 3). Then, using the same technique as in the proof of Lemma 4.4, with C_1, C_2, C_3 playing the role of a, b, c, we show that Duplicator wins the EF^+ game on u, v with $\log(N)$ rounds. Therefore L_M is not FO^+-definable, by Corollary 3.8.

If M mortal

The converse direction is more difficult: we have to show that if there is a bound n on the length of runs of M from any configuration, then L_M is FO^+-definable. We will again use the EF^+-game and Corollary 3.8: we give an integer m (depending only on n) such that for any $u \in L_M$ and $v \notin L_M$, Spoiler wins $\mathrm{EF}^+_m(u, v)$.

Without loss of generality, consider $u = u_0\#u_1\#\ldots\#u_N$ a word of L_M, where each u_i is in C, and v a word not in L_M. To describe the winning strategy of Spoiler, we will first rule out the problems in "local behaviours": if v contains a factor containing at most 2 symbols # preventing it from belonging to L_M, then Spoiler wins easily in a bounded number of rounds by pointing this local inconsistency in v, that cannot be mirrored in u. The only remaining problem is the "long-term inconsistency" occurring in the previous EF^+ games: a long factor with two conflicting possible interpretations, each being forced

by one of the endpoints. For instance, when dealing with the language K, such long-term inconsistencies were exhibited by words of the form $\binom{a}{b}\binom{b}{c}\binom{c}{a}\binom{a}{b}\dots\binom{b}{c}$, with the first letter being constrained to a and the last one to c. We have to show that contrarily to what happens with the language K, or in the case where M is not mortal, Spoiler can now point out such long-term inconsistencies in a bounded number of rounds.

To do that, let us abstract a configuration word $w \in C$ by its *height* $h(w)$: the length of the run of M starting in the configuration w. Our mortality hypothesis can be rewritten as: the height of any configuration word is at most n. A word $w \in C$ will be abstracted by a single letter $h(w) \in [0, n]$. We saw that if a word w' is of the form $\binom{w_1}{w_2}$, then w_1 and w_2 encode consecutive configurations of M, so their heights must be consecutive integers $i+1$ and i. We will abstract such a word w' by the letter $\binom{i+1}{i}$. This allows us to design an abstracted version of the EF-game, called the *integer game*, where letters are integers or pairs of integers, and with special rules designed to reflect the constraints of the original EF^+ game on (u, v). The integer game makes explicit the core combinatorial argument making use of the mortality hypothesis. We show that Spoiler wins this integer game in $2n$ rounds. We finally conclude by lifting this strategy to the original EF^+-game.

This ends the proof sketch, and we now go to the more detailed proof.

6.2. **The Turing Machine Mortality problem.** We will start by describing the problem we will reduce from, called Turing Machine (TM) Mortality.

The TM Mortality problem asks, given a deterministic TM M, whether there exists a bound $n \in \mathbb{N}$ such that from any finite configuration (state of the machine, position on the tape, and content of the tape), the machine halts in at most n steps. We say that M is *mortal* if such an n exists.

Theorem 6.2 [Hoo66]**.** *The TM Mortality problem is undecidable.*

Remark 6.3. The standard mortality problem as formulated in [Hoo66] does not ask for a uniform bound on the halting time, and allows for infinite configurations, but it is well-known that the two formulations are equivalent using a compactness argument. Indeed, if for all $n \in \mathbb{N}$, the TM has a run of length at least n from some configuration C_n, then we can find a configuration C that is a limit of a subsequence of $(C_n)_{n\in\mathbb{N}}$, so that M has an infinite run from C.

Notice that the initial and final states of M play no role here, so we will omit them in the description of M. Indeed, we can assume that M halts whenever there is no transition from the current configuration.

Let $M = (\Gamma, Q, \Delta)$ be a deterministic TM, where Γ is the alphabet of M, Q its set of states, and $\Delta \subseteq Q \times \Gamma \times Q \times \Gamma \times \{\leftarrow, \rightarrow\}$ its (deterministic) transition table.

We will also assume without loss of generality that Q is partitioned into Q_1, Q_2, Q_3, and that all possible successors of a state in Q_1 (resp. Q_2, Q_3) are in Q_2 (resp. Q_3, Q_1). Remark that if M is not of this shape, it suffices to make three copies Q_1, Q_2, Q_3 of its state space, and have each transition change copy according to the 1-2-3 order given above. This transformation does not change the mortality of M.
We will say that p has *type* i if $p \in Q_i$. The *successor type* of 1 (resp. 2, 3) is 2 (resp. 3, 1).

Our goal is now to start from an instance M of TM Mortality, and define a regular language L_M such that L_M is FO^+-definable if and only if M is mortal.

6.3. The base language L_{base}.

The base alphabet

We define first a *base alphabet* A_{base}. Words over this alphabet will be used to encode configurations of the TM M.

$$A_{base} = \Gamma \cup (\Delta \times \Gamma) \cup (\Gamma \times \Delta) \cup (\Delta \times \Gamma \times \Delta) \cup (Q \times \Gamma) \cup \{\#\}.$$

We will note a_δ (resp. $a^{\delta'}, a_\delta^{\delta'}$) the letters from $\Delta \times \Gamma$ (resp. $\Gamma \times \Delta, \Delta \times \Gamma \times \Delta$), and $[q.a]$ letters of $Q \times \Gamma$.

The letter $[q.a]$ is used to encode the position of the reading head, $q \in Q$ being the current state of the machine, and $a \in \Gamma$ the letter it is reading.

A letter a_δ will be used to encode a position of the tape that the reading head just left, via a transition δ writing an a on this position. A letter $a^{\delta'}$ will be used for a position of the tape containing a, and that the reading head is about to enter via a transition δ'. We use $a_\delta^{\delta'}$ if both are simultaneously true, i.e. the reading head is coming back to the position it just visited. Finally, the letter # is used as separator between different configurations.

Configuration words

The encoding of a configuration of M is therefore a word of the form (for example):

$$a_1 a_2 \dots (a_{i-1})^{\delta'} [q.a_i] (a_{i+1})_\delta \dots a_n.$$

The letter $(a_{i+1})_\delta$ indicates that the reading head came from the right via a transition $\delta = (_, _, q, a_{i+1}, \leftarrow)$ (where $_$ is a placeholder for an unknown element). The letter $(a_{i-1})^{\delta'}$ indicates that it will go in the next step to the left via a transition $\delta' = (q, a_i, _, _, \leftarrow)$.

A word $u \in (A_{base})^*$ is a *configuration word* if it encodes a configuration of M with no inconsistency. More formally, u is a configuration word if u contains no #, exactly one letter from $Q \times \Gamma$ (the reading head), and either one a_δ and one $b^{\delta'}$ located on each side of the head, or just one letter $a_\delta^{\delta'}$ adjacent to the head. Moreover, the labels δ and δ' both have to be coherent with the current content of the tape.

Remark 6.4. Because we ask these δ and δ' labellings to be present, configuration words only encode TM configurations that have a predecessor and a successor configuration.

The *type* of a configuration word is simply the type in $\{1, 2, 3\}$ of the unique state it contains.

Let us call $C \subseteq (A_{base})^*$ the language of configuration words. This language C is partitioned into C_1, C_2, C_3 according to the type of the configuration word. It is straightforward to verify that each C_i is an FO-definable language.

We can now define the language L_{base}. The basic idea is that we want L_{base} to be $(C_1\#C_2\#C_3\#)^*$, but in order to avoid unnecessary bookkeeping later in the proof, we do not want to care about the endpoints being C_1 and C_3. Let us also drop the last # which is useless as a separator, and assume that C_1 appears at least once, just for the sake of simplifying the final expression. This gives for L_{base} the expression:

$$(\varepsilon + C_3\# + C_2\#C_3\#)(C_1\#C_2\#C_3\#)^*(C_1 + C_1\#C_2 + C_1\#C_2\#C_3).$$

Notice that L_{base} cannot verify that the sequence is an actual run of M, since it just controls that the immediate neighbourhood of the reading head is valid, and that the types succeed each other according to the 1-2-3 cycle. The tape can be arbitrarily changed from one configuration word to the next.

6.4. **The alphabet A.** We now define another alphabet A_{amb} (amb for ambiguous), consisting of some unordered pairs of distinct letters from A_{base}. An unordered pair $\{a, b\}$ is in A_{amb} if a can be replaced with b in the encodings of two successive configurations of M of the same length (with surrounding letters changing as well). Thus, let A_{amb} be the following set of unordered pairs (we note the "predecessor" element first to facilitate the reading):

- $\{a_\delta, a\}$, $a \in \Gamma$, $\delta \in \Delta$
- $\{a, a^{\delta'}\}$, $a \in \Gamma$, $\delta' \in \Delta$
- $\{a^{\delta'}, [q.a]\}$, $\delta' = (_, _, q, _, _) \in \Delta$
- $\{a_\delta^{\delta'}, [q.a]\}$, $\delta = (_, _, p, a, d) \in \Delta$, $\delta' = (p, _, q, _, -d) \in \Delta$
- $\{[p.a], b_\delta\}$, $\delta = (p, a, _, b, _) \in \Delta$
- $\{[p.a], b_\delta^{\delta'}\}$, $\delta = (p, a, q, b, d) \in \Delta$, $\delta' = (q, _, _, _, -d) \in \Delta$

where $_$ stands for an arbitrary element, and $-d$ is the direction opposite to d.

Notice that all letters of A_{amb} have a clear "predecessor" element: even the possible ambiguity regarding letters $a_\delta^{\delta'}$ are resolved thanks to the type constraint on transitions of M. For readability, we will use the notation $\binom{a}{b}$ instead of $\{a, b\}$, where the upper letter is the predecessor element.

We can now define the alphabet $A = A_{base} \cup A_{amb}$, partially ordered by $a <_A b$ if $a \in A_{base}, b \in A_{amb}, a \in b$. The order $\leq_A$ is the reflexive closure of $<_A$. For now we use the general formalism of ordered alphabet for simplicity. We will later describe in Remark 6.25 how the construction is easily modified to fit in the powerset alphabet framework.

6.5. **Superposing configuration words.** We will see that thanks to the definition of the alphabet A_{amb}, two distinct configurations can be "superposed", i.e. can be written simultaneously with letters of A including letters of A_{amb}, if and only if one follows from the other by a valid transition of M.

Lemma 6.5. *If $u_1, u_2 \in C$ encode two successive configurations of the same length, then there exists $v \in A^*$ such that $u_1 \leq_A v$ and $u_2 \leq_A v$.*

Proof. It suffices to take the letters in v to be the union of letters in u_1, u_2 when these letters differ. For instance if $u_1 = aab^{\delta'}[p.a]c_\delta c$ and $u_2 = aa^{\delta''}[q.b]d_{\delta'}cc$, then $v = a\binom{a}{a^{\delta''}}\binom{b^{\delta'}}{[q.b]}\binom{[p.a]}{d_{\delta'}}\binom{c_\delta}{c}c$. □

Lemma 6.6. *Let $u_1, u_2 \in C$, and assume that there exists $v \in A^*$ satisfying $u_1 \leq_A v$ and $u_2 \leq_A v$. Then either $u_1 = u_2$, or one is the successor configuration of the other.*

Proof. Let $[p.a]$ and $[q.b]$ be the reading heads in u_1 and u_2, in positions i and j respectively.

If $i = j$, let us consider the letter $v[i]$, we know that $[p.a] \leq_A v[i]$ and $[q.b] \leq_A v[i]$. Since no letter of the form $\{[p.a], [q.b]\}$ exists in A_{amb}, we must have $[p.a] = [q.b]$. Let $\delta' = (p, a, p', a', d)$ be the transition of M from $[p.a]$. Let $\lambda = -1$ if $d = \leftarrow$ and $\lambda = 1$ if $d = \rightarrow$. By definition of C, we must have letters b_1, b_2 such that $u_1[i + \lambda] = b_1^{\delta'}$ and $u_2[i + \lambda] = b_2^{\delta'}$, with possible additional δ subscripts. As before, there is no letter $\{b_1^{\delta'}, b_2^{\delta'}\}$ in A_{amb}, even with optional additional δ subscripts, so $u_1[i + \lambda] = u_2[i + \lambda]$. Finally, either both $u_1[i - \lambda]$ and $u_2[i - \lambda]$ are letters from Γ (if the δ subscript is present $u_1[i + \lambda] = u_2[i + \lambda]$), or are letters with a δ subscript. In both cases, again by definition of A_{amb}, we must have $u_1[i - \lambda] = u_2[i - \lambda]$. The rest of the words u_1, u_2 outside of these three positions $\{i_1, i, i + 1\}$ are forced to be letters of Γ by definition of C, so again they cannot vary between u_1 and u_2, since A_{amb} does not contain letters $\{a, b\}$ with $a, b \in \Gamma$. We can conclude that $u_1 = u_2$.

Consider now the case where $i \neq j$. Assume without loss of generality that p is of type 1 (no type plays a particular role). Let δ_1 be the transition from $[p.a]$, of type $1 \to 2$. Let us called *enriched letter* a letter of the form a_δ, $a^{\delta'}$, or $a_\delta^{\delta'}$. By definition of A_{amb}, both $u_2[i]$ and $u_1[j]$ must be enriched letters. By definition of C, it means $u_1[j]$ is just next to $u_1[i]$ where the reading head is, say without loss of generality $j = i + 1$. So we have $v[i] = \{[p.a], u_2[i]\} \in A_{amb}$, and as we saw, $u_2[i]$ is either predecessor or successor to $[p.a]$ in this case. Let us assume for now that $u_2[i]$ is successor to $[p.a]$. This means it has a δ_1 subscript. Since $u_2 \in C$, this forces the state q to be of type 2, the target type of δ_1. Consider now the enriched letter $u_1[j]$, which is such that $\{[q.b], u_1[j]\} \in A_{amb}$. Either $u_1[j]$ has a superscript δ' with target q, or a subscript δ_2 with source p. This latter case is not possible, as from the fact that $u_1 \in C$, it would force p to be of type 3, the target type of δ_2. So we have indeed that both $u_1[i]$ and $u_1[j]$ are the predecessors of $u_2[i]$ and $u_2[j]$ in the letters $v[i], v[j]$ of A_{amb} respectively. It is straightforward to verify that this implies that u_1 is the predecessor configuration of u_2: the only discrepancies allowed between u_1 and u_2 outside of positions i, j are of the form $\{a_\delta, a\}$ and $\{a, a^{\delta'}\}$ and do not influence the underlying letter of Γ. Similarly, in the other case, where $u_2[i]$ is predecessor to $[p.a]$ in the letter $v[i]$, we obtain that u_1 is the successor configuration to u_2. □

Lemma 6.7. *It is impossible to have three distinct words $u_1, u_2, u_3 \in C$ and $v \in A^*$ such that for all $i \in \{1, 2, 3\}$, $u_i \leq_A v$.*

Proof. By Lemma 6.6, any pair from $\{u_1, u_2, u_3\}$ must encode two consecutive configurations of M. However, since the reading head must move at each step, from u_1 to u_2 and from u_2 to u_3, this means the reading head moves either 0 or 2 positions between u_1 and u_3, which yields a contradiction. □

6.6. **The language L_M.** We finally define L_M to be the monotone closure of L_{base} with respect to $\leq_A$, so that L_M can contain letters from A_{amb}.

By Lemma 2.4, since L_M is the monotone closure of a regular language, it is regular (and monotone).

As a side remark, we can observe the following:

Remark 6.8. L_M is FO-definable. Since it is not crucial to the following, we only give here a rough intuition on why L_M is FO-definable. The language K from Section 4 can be seen as an abstraction of L_M, with a, b, c playing the role of C_1, C_2, C_3 respectively. In this light, and since C_1, C_2, C_3 are all FO-definable, we can use the fact that the language K is FO-definable as well, by Lemma 4.3, to obtain an FO-formula for L_M. We also need Lemma 6.7 to guarantee that the equivalent of the letter $\top$ from K never appears.

We will now prove in the next sections that L_M is FO^+-definable if and only if M is mortal, using Corollary 3.8.

6.7. **M not mortal $\implies$ L_M not FO^+-definable.** Let $n \in \mathbb{N}$, we aim to build $(u, v) \in L_M \times \overline{L_M}$ such that $u \preceq_n v$.

There is a configuration from which M has a run of length $N + 3$, with $N = 2^{n+1} + 1$. Let $u = u_0 \# u_1 \# \ldots \# u_N$ be an encoding of this run where each $u_i \in C$, and where we omitted the first and last configurations of the run, which may not be representable in C by

Remark 6.4. Here all the u_i's are of the same length, which is the size of the tape needed for this run.

By Lemma 6.5, for each $i \in [0, N-1]$, there exists $v_i \in A^*$ such that $u_i \leq_A v_i$ and $u_{i+1} \leq_A v_i$.

We build $v = u_0\#v_1\#\ldots\#v_{N-2}\#u_N$. Notice that $v \notin L_M$, because the types of u_0 and u_N forces them to be separated by $N-1 \mod 3$ configurations as in u, but in v they are separated by $N-2 \mod 3$ configurations.

We describe a strategy for Duplicator witnessing $u \preceq_n v$. It is a simple adaptation from the proof of Lemma 4.4, so we will just sketch the idea.

Let us consider that initially, there is a pair of initial (resp. final) tokens at the beginning (resp. end) of u, v. That is, we choose to start by Spoiler playing his first two moves at the first and last position of u, answered correctly by Duplicator at the first and last position of v (otherwise Spoiler easily wins in one move). We will consider that the initial tokens are "blue", and the final ones are "yellow". In the following, a pair of corresponding tokens in u, v will be blue (resp. yellow) if they are at the same distance to the beginning (resp. end) of the word.

When Spoiler plays a token in u_i (resp. v_i), Duplicator will look at the color of the closest token in u_i, (resp. v_i), and answer with a token of the same color, i.e. by playing in v_i (resp. u_i) for blue, and in v_{i-1} (resp. u_{i+1}) for yellow. Of course, the same strategy applies to tokens played on # positions, and an arbitrary choice can be made for positions to both colors.

This strategy preserves the following invariant: after k rounds, the number of # between the last blue token and the first yellow token on the same word (u or v) is at least 2^{n-k}. This invariant guarantees that Duplicator wins the n-round game, since this gap will never be empty.

6.8. M **mortal** $\implies$ L_M **FO$^+$-definable.** Let M be a mortal TM, and n be the length of a maximal run of M, starting from any configuration.

We will show that L_M is FO$^+$-definable, by giving a strategy for Spoiler in $EF^+_{f(n)}(u, v)$ for any $(u, v) \in L_M \times \overline{L_M}$, where the number of rounds $f(n)$ depends only on n, and not on u, v.

Let $(u, v) \in L_M \times \overline{L_M}$. Without loss of generality we can assume that $u \in L_{base}$. This is because there exists $u' \in L_{base}$ with $u' \leq_A u$, and we can consider the pair (u', v) instead of (u, v). Indeed, if Spoiler wins on (u', v), then the same strategy is winning on (u, v), where his winning condition only gets easier.

Thus we can write $u = u_0\#u_1\#\ldots\#u_N$, where each u_i is in C. Let us also write $v = v_0\#v_1\#\ldots\#v_T$, where each v_i does not contain #. Let us emphasize that no assumption is made on N and T, they can be any integers.

We will now describe a strategy for Spoiler in $EF^+(u, v)$, that is winning in a number $f(n)$ of rounds only depending on n.

Ruling out local inconsistencies

As explained in the proof scheme of Section 6.1, we will first show that if v presents *local inconsistencies* (that we define here formally via the notion of forbidden local factor), Spoiler can point them out in a bounded number of moves.

Definition 6.9. Let us call *local factor* a factor containing at most two symbols #. A local factor is *forbidden* if it is not a factor of any word in L_M.

Lemma 6.10. *If v contains a forbidden local factor, Spoiler can win in a constant number of moves (at most 5).*

Proof. This can be seen by verifying that the language of words containing no forbidden local factors is FO^+-definable, with a formula φ_{loc} using at most 5 nested quantifiers. We sketch here how such a formula φ_{loc} can be built.

Notice that formulas of FO^+ can use the letter # either positively or negatively, since it is not comparable with any other letter. If $(p, a) \in Q \times \Gamma$, we define $S(p, a) := \{(\sigma, \tau) \in (A_{base})^2 \mid \sigma[p.a]\tau \in C\}$ as the possible neighbourhoods of $[p.a]$ in the base alphabet. Let us define an FO^+-formula $\varphi_C(x)$ with free variable x, to be a formula that verifies that the maximal #-free factor containing position x is in $C^\uparrow$, and that this is witnessed by the reading head at position x. This formula will verify that x contains some $[p.a]$, that the immediate neighbourhood of x is compatible with $[p.a]$, via a formula $\bigvee_{(\sigma,\tau)\in S(p,a)} \sigma^\uparrow(x-1) \wedge \tau^\uparrow(x+1)$, and that all other letters (not on positions $\{x-1, x, x+1\}$) between the neighbouring # symbols are in $\Gamma^\uparrow$.

We now give a description of the formula φ_{loc}. The formula will state that for all positions $x < y$ of successive # symbols (i.e. with no # between them), there must be positions $i_1 < x < i_2 < y < i_3$, with only two # symbols in $[i_1, i_3]$, such that $\varphi_C(i_1) \wedge \varphi_C(i_2) \wedge \varphi_C(i_3)$. Additionally, the types of the states in i_1, i_2, i_3, must be respectively either 1-2-3, 2-3-1, or 3-1-2. □

From now on, we will therefore assume that v does not contain forbidden local factors.

Finding long-term inconsistencies

We will see how the only remaining cause for v not belonging to L_M is what we called *long-term inconsistencies* in the proof scheme of Section 6.1. We will formalize this with the notion of non-coherent maximal ambiguous factor.

Let us start with an auxiliary definition.

Definition 6.11. A factor v_i of v is *compatible* with type $j \in \{1, 2, 3\}$ if there exists $u' \in C_j$ with $u' \leq_A v_i$. The *set-type* of v_i is $\{j \mid v_i \text{ is compatible with } j\}$.

By Lemma 6.7, each v_i is compatible with at most 2 distinct types in $\{1, 2, 3\}$. If v_i is compatible with 2 types, then one is the predecessor (resp. successor) of the other in the 1-2-3 cycle order, and we call it the *first type* (resp. *second type*) of v_i. We will consider that v_0 (resp. v_T) is only compatible with $type(u_0)$ (resp. $type(u_N)$). Indeed, if Duplicator matches v_0 to a word u_i with $i \neq 0$, Spoiler can win the game in the next round, by choosing a # position before u_i (and same argument for v_T).

Definition 6.12. A factor of the form $v_i\#v_{i+1}\#\ldots\#v_j$ of v is called *ambiguous* if each v_k for $k = i, \ldots, j$ is compatible with two types, and the set-types succeed each other in the cycle order $\{1, 2\} \to \{2, 3\} \to \{3, 1\}$. For instance if the set-type of v_i is $\{3, 1\}$, then v_{i+1} must have set-type $\{1, 2\}$. An ambiguous factor is *maximal* if it is not contained in a strictly larger ambiguous factor.

Definition 6.13. A factor v_i of v is called an *anchor* if either $i = 0, i = T$ or if $v_{i-1}\#v_i\#v_{i+1}$ is not ambiguous.

If v_i is an anchor, we can uniquely define its *anchor type*. It is simply its type if $i = 0$ or T, and otherwise since $v_{i-1}\#v_i\#v_{i+1}$ is not ambiguous, we define the anchor type of v_i to be the only possible type for v_i that does not create an incoherence with its two neighbours. Notice that such a type exists, since we assumed v does not contain forbidden local factors.

Example 6.14. Assume v_5 has set-type $\{2,3\}$, v_6 has set-type $\{3,1\}$, and v_7 has set-type $\{2,3\}$. Then v_6 is an anchor, and its anchor type is 1. The type 3 is indeed impossible for v_6, since its successor type 1 is not in the set-type of v_7.

Notice that if Duplicator maps an anchor v_i to a word u_j such that $type(u_j)$ is not the anchor type of v_i, then Spoiler can win in at most 5 moves, by pointing to a contradiction with the immediate neighbourhood of v_i.

Definition 6.15. Let $v_i\#v_{i+1}\#\ldots\#v_j$ be a maximal ambiguous factor. It is called *coherent* if $v_{i-1}\#v_i\#\ldots\#v_{j+1}$ is an infix of L_M.

Such a coherent factor will be witnessed by the anchor types of v_{i-1} and v_{j+1}: this means that $v_i\#v_{i+1}\#\ldots\#v_j$ is coherent if the anchor types at v_{i-1}, v_{j+1} are either both concatenable with the first type of both v_i, v_j, or are both concatenable with their second type. Here by "concatenable", we mean to respect the 1-2-3 order, for instance type 3 must be followed by type 1.

Example 6.16. Let $w = v_i\#v_{i+1}\#\ldots\#v_j$ be a maximal ambiguous factor, where v_i has set-type $\{1,2\}$ and v_j has set-type $\{2,3\}$. Notice that this implies $j \equiv i+1 \mod 3$. Assume v_{i-1} has anchor type 1, so it is concatenable with the second type of v_i. This means that for w to be coherent, we need v_{j+1} to have anchor type 1, in order to be concatenable with the second type of v_j as well.

Lemma 6.17. *v contains a maximal ambiguous factor w that is not coherent.*

Proof. Assume that all maximal ambiguous factors of v are coherent. Since v does not contain forbidden local factors, we have that the anchor types of two consecutive anchors follow the 1-2-3 order. This means that the anchor types, together with the coherence of maximal ambiguous factors, give us a witness that $v \in L_M$. This witness is a word of L_{base}, obtained by choosing the anchor types on all anchors, and either the first type or the second type uniformly in maximal ambiguous factors, as fixed by the anchors at the extremities. Since we know that $v \notin L_M$, this is a contradiction. □

We are now ready to describe Spoiler's strategy. Spoiler starts by placing two tokens delimiting a maximal ambiguous factor w that is not coherent, as obtained in Lemma 6.17. Because w is not coherent, Duplicator is forced to answer with the first type for one of these tokens, and with the second type for the other: otherwise Spoiler immediately wins by exposing a local inconsistency with the anchors delimiting w.

We will now show how Spoiler can win on such a factor w, starting with these two tokens. For this, we will introduce the notion of height of a configuration word, and the integer game that will abstract the EF^+-game on w. This is where we finally make use of the hypothesis that M is mortal.

Abstracting words by integers

Recall that from the mortality assumption, we have a bound n on the length of runs of M, starting from any configuration.

Definition 6.18. If $u \in C$ is a configuration word, its *height* $h(u)$ is the length of the run starting in u, and not going outside of the tape specified in u.

If $u \in C$, let us also define its n-approximation $\alpha_n(u)$ as the maximal word in $(A_{base})^{\leq n} \cdot (Q \times \Gamma) \cdot (A_{base})^{\leq n}$ that is an infix on u. That is, we remove letters whose distance to the reading head is bigger than n.

Here are a few properties of the height:

Lemma 6.19. *For all $u \in C$, the following hold:*

- $0 \leq h(u) < n$.
- *For all $x, y \in \Gamma^*$, we have $h(xuy) \geq h(u)$.*
- $h(u) = h(\alpha_n(u))$.
- *If $v \in C$ is the successor configuration of $u \in C$, then $h(v) = h(u) - 1$.*

Proof. The first item is a consequence of the fact that M is mortal with bound n. Notice that the inequalities are strict because of Remark 6.4: words from C must have a predecessor and a successor configuration. The second item comes from the fact that the run of length $h(u)$ starting in u is still possible when adding a context x, y, which is not affected by this run. The third item uses the fact that a run can only visit the n-approximation of u, so the context outside of $\alpha_n(u)$ does not affect the height $h(u)$. The fourth item is a basic consequence of the definition of the height. □

Corollary 6.20. *The height of a configuration word u is an FO^+-definable property, i.e for all $k \in \mathbb{N}$ there exists an FO^+ formula h_k such that h_k accepts a configuration word $u \in C$ if and only if $h(u) = k$.*

Proof. By Lemma 6.19, the formula h_k can simply use a lookup table to verify that $\alpha_n(u)$ is of height k, using a finite disjunction listing possibilities for $\alpha_n(u)$ being of height k. When evaluated on A^*, the formula h_k will accept the monotone closure of configuration words of height k. □

Remark 6.21. We use here the fact that computation is done locally around the reading head to obtain Corollary 6.20. This seems to make Turing Machines more suited to this reduction than e.g. cellular automata, where computation is done in parallel on the whole tape.

Thanks to the height abstraction, we will show that we can focus on playing a special kind of abstracted EF-game.

The integer game

The idea is to abstract a configuration word $u_i \in C$ by its height $h(u_i)$. If u' is the predecessor configuration of u'', and $u', u'' \leq_A v$, we will abstract the word v by $\binom{h(u')}{h(u'')}$.

Let $\Sigma_{base} = [0, n]$ and $\Sigma_{amb} = \{\binom{i}{i-1} \mid 1 \leq i \leq n\}$. Let $\Sigma = \Sigma_{base} \cup \Sigma_{amb}$, ordered by $i \leq_\Sigma \binom{i}{i-1}$ and $i - 1 \leq_\Sigma \binom{i}{i-1}$ for all $\binom{i}{i-1} \in \Sigma_{amb}$.

We define the n-integer game as follows: It is played on an arena (U, V) with $U \in (\Sigma_{base})^*$ and $V \in (\Sigma_{amb})^*$. If we note i (resp. j) the first (resp. last) letter of U, then the first (resp. last) letter of V is $\binom{i}{i-1}$ (resp. $\binom{j+1}{j}$).

The rest of the rules is very close to those of $EF^+(U, V)$: in each round, Spoiler plays a token in U or V, Duplicator has to answer with a token in the other word, while maintaining the order between tokens, and the constraint that the label of a token in U is $\leq_\Sigma$-smaller

than the label of its counterpart in V. We add an additional *neighbouring constraint* for Duplicator: consecutive tokens in one word must be related to consecutive tokens in the other, and in this case, if two tokens of V are in consecutive positions labelled $\binom{i}{i-1}\binom{j}{j-1}$, the corresponding tokens in U must be either labelled i, j or $i-1, j-1$. A mix $i, j-1$ or $i-1, j$ is not allowed.

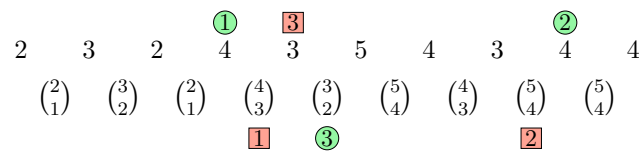


Figure 8: A position of the integer game.

Lemma 6.22. *For all $n \in \mathbb{N}$, Spoiler can win any n-integer game in $2n$ rounds.*

Proof. Let (u, v) be an arena for an n-integer game. We proceed by induction on n.

For $n = 1$, the constraints on the game forces $u \in 1(0+1)^*0$ and $v \in \binom{1}{0}^*$.

We can have Spoiler play on the last occurrence of 1 in u, and on the successor position labelled 0. Duplicator cannot respond to these two moves while respecting the neighbouring constraint, so Spoiler wins in 2 moves.

Assume now that for some $n \geq 1$, Spoiler wins any n-integer game in $2n$ moves, and consider an $(n+1)$-integer game arena (u, v). If the letters $n+1$ and $\binom{n+1}{n}$ do not appear in u, v respectively, then Spoiler can win in $2n$ moves by induction hypothesis.

If the letter $n+1$ does not appear in u, then let y be the first position labelled $\binom{n+1}{n}$ in v. By definition of the integer game y cannot be the first position of v, otherwise u should start with $n+1$. We will choose position y in v for the first move of Spoiler, let x be the position in u answered by Duplicator, we have $u[x] = n$. We can assume that x is not the first position of u, otherwise Spoiler can win in the next move. If Spoiler were to play $x-1$ in u, with $u[x-1] = i$, by the neighbouring constraint Duplicator would be forced to answer $y-1$ in v, with label $\binom{i+1}{i}$. This shows that the words $u[..x-1]$ and $v[..y-1]$ form a correct n-integer arena, as the integer $n+1$ is not present anymore, and all other constraints are respected. Therefore, Spoiler can win by playing $2n$ moves in these prefixes. This gives a total of $2n+1$ moves in the original $(n+1)$-integer game.

Finally, if the letter $n+1$ does appear in u, Spoiler starts by playing the position x in u corresponding to the last occurrence of $n+1$ in u. Duplicator must answer a position y labelled $\binom{n+1}{n}$. Notice that neither x nor y can be a last position, so $u[x+1]$ and $v[y+1]$ are well-defined. As before, using the neighbouring constraint, we know that if $i = u[x+1]$, then $v[y+1] = \binom{i}{i-1}$. Therefore, the words $u[x+1..]$ and $v[y+1..]$ form an $(n+1)$-integer game arena, and moreover the letter $n+1$ does not appear in $u[x+1..]$ (by choice of x). Using the precedent case, we know that Spoiler can win from there in $2n+1$ moves, playing only on $u[x+1..]$ and $v[y+1..]$. This gives a total of $2n+2$ moves in the original $(n+1)$-integer game, thereby completing the induction proof. □

Remark 6.23. Lemma 6.22 still holds if the definition of n-integer game is generalized to include the symmetric case where, if we note i, j the first and last letters of U respectively, V starts with $\binom{i+1}{i}$ and ends with $\binom{j}{j-1}$. Indeed, it suffices to consider the mirrored images of U and V to show that Spoiler wins in the same amount of rounds.

We can now see how to use Lemma 6.22 (together with Remark 6.23) to conclude the proof.

Lifting the strategy to the original EF^+-game

Let us go back to the EF^+-game on u, v, where Spoiler has placed two tokens delimiting a non-coherent maximal ambiguous factor w in v.

Spoiler can now play only between these existing tokens, and import the strategy from the integer game, by abstracting each word u_i by its height and each word v_j by $\binom{h(u')}{h(u'')}$, where $u', u'' \in C$ are such that $u' \leq_A v_j$, $u'' \leq_A v_j$, and $h(u') = 1 + h(u'')$. Each factor of u_i, v_j delimited by # corresponds to a single position in the abstracted integer game. Spoiler can for instance mimic a move of the integer game by playing in the first position of the corresponding factor in u_i or v_j, i.e. just after a #-labelled position.

Lemma 6.24. *If Duplicator does not comply with the rules of the integer game, then Spoiler can win in at most* $\log n$ *rounds.*

Proof. Assume u_i is matched to v_j, but there is no $u' \leq_A v_j$ such that $h(u_i) = h(u')$. It means that $\alpha_n(u)$ is not compatible with v, and this can be exploited by Spoiler using $\log n$ rounds (with a dichotomy strategy, or n rounds with a naive strategy). Thus by Lemma 6.19, Spoiler can enforce the basic rule of the integer game, stating that if integer t is matched to $\binom{s+1}{s}$, then $t = s + 1$ or $t = s$. Using the correspondence between EF^+-games and FO^+-definability, this property can also be seen via Corollary 6.20.

If neighbours are matched with non-neighbours, then it suffices for Spoiler to point the two # positions between the non-neighbours, that cannot be matched in the other word, so he wins in 2 moves. We show that the rest of the neighbourhood rule is also enforced. Assume $u_i\#u_{i+1}$ is matched to $v_j\#v_{j+1}$. Assume $\mathit{type}(u_i)$ is the first (resp. second) type of v_j while $\mathit{type}(u_{i+1})$ is the second (resp. first) type of v_{j+1}. By definition of L_M, $\mathit{type}(u_{i+1})$ must be the successor type of $\mathit{type}(u_i)$, for instance without loss of generality, $\mathit{type}(u_i) = 1$ and $\mathit{type}(u_{i+1}) = 2$. Then, the set-type of v_i is $\{1, 2\}$ (resp. $\{3, 1\}$) and the set-type of v_{j+1} is $\{1, 2\}$ (resp. $\{2, 3\}$). This contradicts the fact that $v_i\#v_{i+1}$ is part of an ambiguous factor, as set-types should follow each other in the order $\{1, 2\}$-$\{2, 3\}$-$\{3, 1\}$. □

Combining these arguments and by Lemma 6.22 and Remark 6.23, we obtain that following this strategy, Spoiler will win in at most $f(n) = 2 + 2n + \log n + 5$ rounds, by either winning the related n-integer game, or exploiting local inconsistencies if Duplicator fails to comply to the rules of the n-integer games. The $+5$ comes from Lemma 6.10 about forbidden local factors.

Using Corollary 3.8, we obtain that L_M is FO^+-definable, with a formula of quantifier rank at most $f(n)$.

Remark 6.25. The alphabet A can be turned into a powerset alphabet, by adding all subsets of A_{base} absent from A_{amb}, rejecting any word containing $\emptyset$ but no new non-empty subset, and accepting any word containing a new non-empty subset. This shows that this undecidability result still holds in the special case of powerset alphabets.

This concludes the proof of Theorem 6.1, up to the proof of Lemma 6.22 which is done in the next section.

Conclusion. We believe this paper gives an example of fruitful interaction between automata theory and model theory. Indeed, a classical result of model theory, the failure of Lyndon's theorem on finite structures, has been greatly simplified by using the toolbox of regular languages. Moreover, our investigation of this question via regular languages also brings a new result: the failure of Lyndon's theorem on finite graphs. Conversely, this question coming from model theory, when considered on regular languages, yields the first (to our knowledge) natural fragment of regular languages with undecidable membership problem, and opens new techniques for proving undecidability of expressibility in positive logics. We hope that the tools developed in this paper can be further used in both fields, and that this will encourage more interactions of this form in the future.

Some questions remain open: for instance what happens for partially ordered alphabets that do not embed the one used in our counter-example language? Can a simpler counter-example be found, or is this structure with two layers of three elements each necessary?

We made the signatures needed for our constructions explicit, in several formulations of Lyndon's theorem, and compared them to the ones used in [AG87, Sto95, Ros95]. It can be investigated whether further improvements on these signatures are possible.

We can also investigate the expressive power of fragments of FO^+ such as the two-variable fragment, or fragments defined by bounding the quantifier rank.

In the short term, we are interested in extending these techniques to the framework of cost functions, see [Kup14, Kup], and to other extensions of regular languages.

Acknowledgements. I am grateful to Thomas Colcombet for bringing this topic to my attention, and in particular for asking the question $\mathrm{FO}^+ \stackrel{?}{=}$ monotone FO, as well as for many interesting exchanges. Thanks also to Amina Doumane and Sam Van Gool for helpful discussions, and to the anonymous reviewers, as well as Anupam Das, Natacha Portier, and Bruno Pasqualotto Cavalar for their comments on earlier versions of this document.

References

[AG87] Miklos Ajtai and Yuri Gurevich. Monotone versus positive. *J. ACM*, 34(4):1004–1015, October 1987.

[AG97] Natasha Alechina and Yuri Gurevich. *Syntax vs. Semantics on Finite Structures*, pages 14–33. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.

[Boj04] Mikołaj Bojańczyk. A bounding quantifier. In Jerzy Marcinkowski and Andrzej Tarlecki, editors, *Computer Science Logic*, pages 41–55, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[Col11] Thomas Colcombet. Green's relations and their use in automata theory. In *Language and Automata Theory and Applications - 5th International Conference, LATA 2011, Tarragona, Spain, May 26-31, 2011. Proceedings*, volume 6638 of *Lecture Notes in Computer Science*, pages 1–21. Springer, 2011.

[Col12] Thomas Colcombet. Regular cost functions, part i: Logic and algebra over words. volume 9, 12 2012.

[Col13] Thomas Colcombet. Magnitude monadic logic over words and the use of relative internal set theory. In *2013 28th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 123–123, 2013.

[DG08] Volker Diekert and Paul Gastin. First-order definable languages. In *Logic and Automata: History and Perspectives, Texts in Logic and Games*, pages 261–306. Amsterdam University Press, 2008.

[FSS81] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. In *22nd Annual Symposium on Foundations of Computer Science (sfcs 1981)*, pages 260–270, 1981.

[GS92] Michelangelo Grigni and Michael Sipser. Monotone complexity. In *Poceedings of the London Mathematical Society Symposium on Boolean Function Complexity*, page 57–75, USA, 1992. Cambridge University Press.

[Hoo66] Philip K. Hooper. The undecidability of the turing machine immortality problem. *Journal of Symbolic Logic*, 31(2):219–234, 1966.

[Kam68] Hans W. Kamp. *Tense Logic and the Theory of Linear Order*. Phd thesis, University of Warsaw, 1968.

[Kup] Denis Kuperberg. Erratum for [Kup14]. URL: `perso.ens-lyon.fr/denis.kuperberg/papers/Erratum.pdf`.

[Kup14] Denis Kuperberg. Linear temporal logic for regular cost functions. *Logical Methods in Computer Science*, 10(1), 2014.

[Kup21] Denis Kuperberg. Positive first-order logic on words. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*, pages 1–13. IEEE, 2021.

[KVB12] Denis Kuperberg and Michael Vanden Boom. On the expressive power of cost logics over infinite words. In *Automata, Languages, and Programming*, pages 287–298, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[Lib04] Leonid Libkin. *Elements of Finite Model Theory*. Springer, August 2004.

[LSS96] C. Lautemann, T. Schwentick, and I. A. Stewart. On positive p. In *2012 IEEE 27th Conference on Computational Complexity*, page 162. IEEE Computer Society, 1996.

[Lyn59] Roger C. Lyndon. Properties preserved under homomorphism. *Pacific J. Math.*, 9(1):143–154, 1959. URL: `https://projecteuclid.org:443/euclid.pjm/1103039459`.

[MP71] Robert McNaughton and Seymour A. Papert. *Counter-Free Automata (M.I.T. Research Monograph No. 65)*. The MIT Press, 1971.

[PST89] Jean-Eric Pin, Howard Straubing, and Denis Therien. New results on the generalized star-height problem. volume 349, pages 458–467, 02 1989.

[PZ19] Thomas Place and Marc Zeitoun. Going higher in first-order quantifier alternation hierarchies on words. *J. ACM*, 66(2), March 2019.

[Ros95] *Finite model theory and finite variable logic*. Phd thesis, University of Pennsylvania, 1995.

[Ros08] Benjamin Rossman. Homomorphism preservation theorems. *J. ACM*, 55, 07 2008.

[Sch65] M.P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8(2):190 – 194, 1965. URL: `http://www.sciencedirect.com/science/article/pii/S0019995865901087`.

[Ste94] Iain A. Stewart. Logical Description of Monotone NP Problems. *Journal of Logic and Computation*, 4(4):337–357, 1994.

[Sto95] Alexei P. Stolboushkin. Finitely monotone properties. In *LICS, San Diego, California, USA, June 26-29, 1995*, pages 324–330. IEEE Computer Society, 1995.

This work is licensed under the Creative Commons Attribution License. To view a copy of this license, visit `https://creativecommons.org/licenses/by/4.0/` or send a letter to Creative Commons, 171 Second St, Suite 300, San Francisco, CA 94105, USA, or Eisenacher Strasse 2, 10777 Berlin, Germany