University of Warsaw
Faculty of Mathematics, Informatics, and Mechanics

VINCENT MICHIELINI

UNIFORMISATION AND CHOICE QUESTIONS FOR REGULAR LANGUAGES

PhD dissertation

Supervisors

MICHAŁ SKRZYPCZAK

Institute of Informatics
University of Warsaw

NATHAN LHOTE

Laboratoire d'Informatique et Systèmes
Aix-Marseille Université

Author's declaration:
Aware of legal responsibility, I hereby declare that I have written this dissertation myself, and that all the contents of the dissertation have been obtained via legal means.

13th September, 2021 ....

VINCENT MICHIELINI

Supervisors' declaration:
The dissertation is ready to be reviewed.

13th September, 2021 ....

MICHAŁ SKRZYPCZAK

....

NATHAN LHOTE

Keywords: Uniformisation, Choice,
Monadic Second-Order Logic,
Algebraic structures

ACM Subject Classification: F.4.3 Formal Languages
F.4.1 Mathematical Logic

Acknowledgements

My first thanks naturally go to the University of Warsaw and to my two supervisors: Michał Skrzypczak and Nathan Lhote. Michale, thank you so much for having given me the chance to work with you. I will write it very shortly: I could not have had a better supervisor. Your professionalism was a blessing during these four years, and you can be sure that your high competence in so many domains will inspire me for a long time. Nathan, merci beaucoup d'avoir accepté de co-superviser ce travail ces deux années durant : toi et tes nombreuses connaissances avez apporté un regard neuf sur cette thèse au meilleur des moments, et ton implication fut d'une grande aide. I hope that you both enjoyed working with me as much as I enjoyed working with you, and I am very impatient to have new occasions to do so.

I continue by thanking all the scientists I met in the University of Warsaw, whether they are part of the automata team or not: Janusz Schmude (dziękuję bardzo, że zawsze byłeś tam dla mnie, życzę Ci wielu sukcesów w Twoich nowych projektach), Adrien Boiret (à toi aussi, merci beaucoup pour ta présence à un moment où j'en avais vraiment besoin, et à bientôt j'espère !), Théodore Lopez, Alexandre Vigny (merci d'avoir répondu à mes appels à l'aide du mercredi matin !), Amina Doumane, Denis Kuperberg, Mikołaj Bojańczyk, whom I owe my very presence at Warsaw!, Eleonora Anna Romano, Rafał Stefański, Mathieu Mari, Oscar Defrain, Lorenzo Clemente, Szymon Toruńczyk, Sandra Kiefer, Jędrzej Kołodziejski, Wojciech Czerwiński, Bartosz Klin, Mohnish Pattathurajan… Sadly I cannot list all the scientists I had the opportunity to share a discussion with, but please make sure that I keep these precious moments in my mind somewhere. I also wish to thank all the smiling and cheerful employees I met daily in MIMUW: the secretaries, the cleaners…: thank you for making the university a very pleasurable place to work in!

Regarding the thesis in itself, I wish to thank Alexander Rabinovich for his very valuable work on the domain of formal languages: the very core of this thesis dates back to him and his interrogations about uniformisations in First-Order Logic.

My next paragraph will be dedicated to the beautiful city of Warsaw, and all the people I met there, who made me enjoy my stay here. My first thanks will go to the two Polish teachers I had the chance to study with: Magdalena Nędzi and Ania Czaja. You helped me feel at home here in Poland, and for this I will always be grateful to you! Then, thanks to the francophone parish on Rakowiecka and the people I shared masses with through the year: father Michał Kamawalski, father Zbigniew Kubacki, Thibault Mouton, Agathe Indihar, Guillaume Bouëssel du Bourg, Flavien Bruant, Ignace Barek, Pierre Mossang, Laure-Marie Sansguilhem, and many more… I also thank the Taizé community who integrated me in the best way possible, here in another non-exhaustive list: Zofia Utri, Marysia Michnowska, Beata Sylwestrzak, Aleksander (bardzo przepraszam, że nie znam twojego nazwiska…!), Antek Starzyński, Magdalena Brzozowska, Julia Osęka, Яна Кожурно… Besides, for sure I cannot speak of Warsaw without mentioning the francophone group of the Tuesday evening, here again the list cannot be exhaustive: thank you Vincent Moser, Dorian Czapla, Mathieu with one t Tachini, Matthieu with two t's Konarzewski, my friend Чавдар Чернев, Pierre Laurent, Marta Borowska, Simon Haddad… Finally, thank you Cécilia Moutet and Nicola Thomas, for all the nice moments in the capital (and not only!), together with Nathan!

I thank now the École Normale Supérieure de Lyon for having accepted me in its computer science department and provided me many nice classes. In particular, thanks to Damien Pous, who then suggested me the automata team of Warsaw as a destination for my PhD. A brief hello to all the fellow guys of my promotions: Pierre "Dionysus" Ohlmann, Rémi Grünblatt, Titouan Carette, Étienne Moutot, Isao Sauzedde, Grégoire Beaudoire, Laureline Pinault, Eugenia Oshurko…, to my room-mates Mickaël Montessinos (promis, je reviens bientôt à toi et je me remets aux catégories !!) and Arthur Charlet, and, of course, to all the people from the chaplain of Gerland: father Matthieu Thouvenot, father Jean-Baptiste Aubert, my godmother Solène Fisch, of course, my godfather Jean Peytou, my forever rival Garance Pontier (c'est chouné !), Amélie Chardac, David Dumont, Étienne Bigné and Hélène Lorenzini, Rémi and Sabine Nollet, Benoît and Léa Zeringer (vous avez vu : je n'ai pas mis l'accent !), Suzanne Brunner, Bertille Petit, Herménégilde Valentin, Nicolas Pinson, Chloé-Nawel Cacheux, Clément and Marielle Brasseur, Clément and Avril Surun, Éric and Maÿlis Woillez-de Saint Savin, Héloïse Guinard, Maureen Boyard (oui, désolé, je t'ai rangée ici faute de mieux…), Jeanne Prillard (même remarque…, et, tant que j'y suis : 1. e4), Rémi and Marie Pellerin… Vous êtes (mal)heureusement beaucoup trop nombreux pour que je puisse tous vous citer, mais vous savez de toute façon combien vous avez compté pour moi durant toutes ces années ! My last but not least thoughts of the paragraph go to *mon kheyou* Fabrice Mouhartem.

I wish to give a special thank to the several internship supervisors I had during my formation at the ÉNS: Luigi Liquori, Ivan Scagnetto, Furio Honsell, and Matteo Mio (the fact they are all Italian was not on purpose!). I am very grateful to you for having guided me when I was doing my first steps as a researcher!

Thanks also go to my preparatory teachers in Fénelon: Jérôme Nizon, Pierre Bouissou, Francis Dorra, Éric Merle, and Muriel Dujardin. I wish to apologise to the last for my awful level in physics back then. Naturally, it is impossible to even mention Fénelon without thinking about you guys: Chris Berd, Christophe Lucas, Anne-Élisabeth Falq, Franck Djeumou, Eelaman Ketheeswaran, Aurélien Zakowic, Maxence Baichère…

Going even more back in time, I need to thank my high school teachers Édouard Guinet and Nathalie Le Dévéhat, who awoke my hidden but yet not-so-hidden interest in mathematics. Finally, my last scholar thanks go to the late Denis Céleste, and his entertaining logical riddles on Friday mornings, back then when I was ten. Reposez en paix.

Merci à toi, Erlé Crouseilles, tout simplement pour être un aussi bon ami depuis tout ce temps.

Finally, this penultimate paragraph goes to my parents, who always supported me and encouraged me in doing what I am passionate about. I am conscious to have received from you what many people do not. In a more general way, credits have to go to all the members of my large family as well: brother, sister, cousins, uncles, aunts, late grandfathers, grandmothers, and many more…: thank you for everything.

My PhD has been supported by ERC Consolidator Grant LIPA (id. 683080).

Abstract

A *uniformisation* of a binary relation $R \subseteq X \times Y$ assigns to each element x of the domain of R a particular $y \in Y$ such that the pair $\langle x, y \rangle$ is in the relation. The *regular-uniformisation problem* asks if, in the context of words or trees, every regular relation admits a uniformisation that is also regular, *i.e.* definable in *Monadic Second-Order Logic* (MSO).

It is already known that the answer to this question is positive in the context of finite and infinite words. In this thesis, we search for possible generalisations of these results.

First, we study the possibility of uniformising relations of finite words in fragments of *First-Order Logic* (FO), the formalism where one can quantify over positions but not over sets of positions. We rely on algebraic characterisations of these fragments in order to emphasise their limitations in providing uniformisations. We also discuss the decidability of uniformising given regular relations in some of these fragments.

Second, we highlight the strong connection between uniformisations and the expressive power of the considered formalisms. Namely, among varieties of regular languages, the assumption that the given formalism has uniformisation property, already guarantees the full power of all regular languages. The notion of a *variety of languages* is a standard concept which guarantees certain natural closure properties, such as closure under Boolean operations.

Third, we study the possibility to uniformise regular relations of words defined over *finitary domains*: those are the domains that are infinite but admit a particular kind of finite representation. We show that, over these domains, the only obstacle to regular uniformisations is the existence of *non-trivial automorphisms*, meaning bijective functions that preserve the order and that are not the identity function. We also prove that, in this context of finitary domains, the regular-uniformisation problem is equivalent to being able to define in MSO natural objects over the domain, such as choice functions and well orders.

All these results provide a broad perspective on the notion of uniformisation in the realm of subclasses and variants of regular languages.

Streszczenie

Uniformizacja relacji binarnej $R \subseteq X \times Y$ przypisuje, dla każdego elementu x z dziedziny R pewien unikalny element $y \in Y$ w taki sposób by para $\langle x, y \rangle$ należała do relacji R. Problem *regularnej uniformizacji* pyta czy, w kontekście słów lub drzew, każda regularna relacja posiada uniformizację która również jest regularna, czyli definiowalna w monadycznej logice drugiego rzędu (MSO).

Wiadomym jest, że odpowiedź na to pytanie jest pozytywna w przypadku skończonych i nieskończonych słów. W ramach tej rozprawy szukamy możliwych rozszerzeń tych wyników.

Po pierwsze, studiujemy możliwość uniformizacji relacji słów skończonych we fragmentach logiki pierwszego rzędu (FO), formalizmu gdzie możliwa jest kwantyfikacja po pozycjach danej struktury, ale nie po zbiorach takich pozycji. Polegamy przy tym na algebraicznych charakteryzacjach rozważanych fragmentów by unaocznić ich ograniczenia w możliwości definiowania uniformizacji. Dodatkowo badamy problem rozstrzygania uniformizowalności danej regularnej relacji w niektórych rozważanych fragmentach.

Po wtóre, podkreślamy silne związki pomiędzy uniformizacjami a siłą wyrazu rozważanych formalizmów. Dokładniej, wśród rozmaitości języków regularnych, założenie że dany formalizm posiada własność uniformizacji gwarantuje nam pełną siłę wyrazu wszystkich języków regularnych. Użyty tu koncept rozmaitości języków jest standardowym pojęciem gwarantującym pewne naturalne własności domknięcia, jak domknięcie ze względu na operacje boolowskie.

Po trzecie, studiujemy możliwość uniformizacji relacji regularnych na słowach definiowanych nad dziedzinami finitarnymi: dziedzinami które są nieskończone, ale dopuszczają pewien szczególny rodzaj skończonej reprezentacji. Wykazujemy, że nad tymi dziedzinami, jedyną przeszkodą ku istnieniu regularnych unifomizacji jest istnienie nietrywialnych automorfizmów, czyli bijekcji które zachowują porządek, ale nie są identycznościowe. Dodatkowo wykazujemy, że w kontekście dziedzin finitarnych, problem regularnej uniformizowalności jest równoważny możliwości zdefiniowania z ramach MSO pewnych naturalnych obiektów nad daną dziedziną, jak funkcje wyboru, czy dobre porządki.

Wszystkie te wyniki dostarczają szerokiej perspektywy na pojęcie uniformizacji w sferze podklas i wariantów języków regularnych.

Contents

Introduction

The following thesis studies multiple variants of a general question, that asks, given some property, under which hypotheses it is possible to assign in some formalism an individual witness satisfying it. The question under study is related to areas as diverse as constructibility of mathematics, logic, model theory, formal languages and their correlation with algebra, and the theory of linear and well orders. The thesis is based on the following three published articles: [Mic18], [LMS19], and [MS20], and contains some additional related unpublished results.

Motivations and context

Foundations of mathematics have been a widely studied field of research since the end of the 19th century. At that time, many renowned mathematicians, such as Frege, Hilbert, or Russell, aimed at coming up with a logical system that encompasses mathematics in its entirety. Notably, from their works was born *set theory*, the main ground on which the whole mathematics is still formalised nowadays.

At this period, some mathematicians, among them Brouwer, were critical about the then-obtained formalisms, feeling that they allowed some reasonings that were counter-intuitive. As an example, one of the main objections was that said formalisms could be used to prove mathematical statements such as "there exists an object x that satisfies the property $\mathcal{P}$", without having to provide any concrete example of such an object x.

Brouwer's works in the field of mathematical foundations led to the construction of a family of formalisms, known today as *intuitionistic logics*. These formalisms have the property that, in order to prove statements such as the previous one, one *has* to provide a proper witness of such an object x.

The question of constructiveness of mathematics is still a very active research area today, that has applications in numerous fields of mathematics. Indeed, whether it is in algebra, in arithmetic, or in real analysis, when one comes up with some result stating the existence of objects that satisfy interesting properties, one is inclined to ask if there is a way to describe said objects in a comprehensive way.

In this thesis, we focus on the possibility to provide *choice functions* in a constructive manner. In set theory, a choice over a family of disjoint non-empty sets is a function that selects one element from each of the sets.

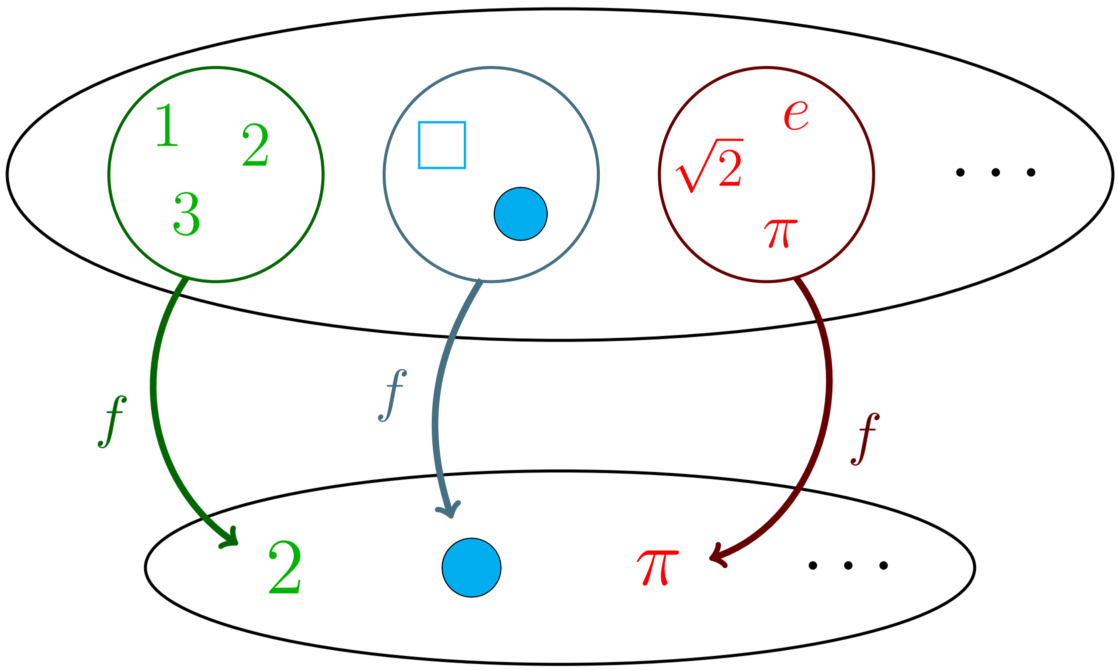


Figure 1: The choice function f selects an element from every set of the family.

The *axiom of choice*, introduced for the first time by Zermelo in [Zer04], states that every such family of subsets admits a choice function. This axiom, considered as intuitive by a majority of mathematicians, has applications in many domains of mathematics. Let us cite a few. In topology, Tychonoff's theorem states that the product of compact sets is also a compact set [Tyc30]. In analysis, Hahn-Banach theorem tells about the possibility to extend a partial linear form into a total one [Ban32]. In linear algebra, it is well known that the axiom of choice is equivalent to the fact that every vector space has a basis (see for instance [Bar13, Lemma 3.1]).

Yet, applying the axiom of choice allows us to obtain mathematical objects and results that are somehow counter-intuitive. For instance, it implies the existence of Vitali sets, *i.e.* sets on which a common notion of measure cannot be defined (see [Vit05]). One could also cite Banach-Tarski paradox, stating that it is possible, via the axiom of choice, to decompose a ball into five parts and put them back together in a different way to obtain two identical copies of the original ball [BT24].

This motivates the search of conditions under which choice functions can be effectively constructed. The main goal of the thesis can be elusively stated as the following:

Problem 0.1. *Let $(A_i)_{i \in I}$ be a family of disjoint non-empty sets. Can we provide a concrete example of a choice function over $(A_i)_{i \in I}$?*

In Chapter XII of his *Introduction to Mathematical Philosophy* [Rus19], Russel gave an illustration of what we could understand by a *concrete example* of a choice function. If the family $(A_i)_{i \in I}$ is an infinite family of pairs of boots, then it is possible to select a unique boot from each pair by simply saying "I choose the right boot of each pair". This choice function is completely comprehensible, as long as we do have a notion of left and right in our mathematical vocabulary.

On the opposite, if the family is an infinite family of pairs of socks, then there is no such natural way of distinguishing the socks of each pair. Of course, if a particular pair is given to us, then we can always proceed to an arbitrary selection of one of the two, but it is impossible to process an infinite amount of such choices. In that case, if we need an infinite selection of socks, then we have to rely on the axiom of choice, that, in a sense, provides this infinite computation for us. However, we cannot know how this infinite computation was made, and we have no clue about the properties satisfied by the obtained infinite set of socks.

Hence, informally speaking, Problem 0.1 asks if we are rather in the case of boots, or in the case of socks.

We will study many variations of Problem 0.1. Among these variations, we mainly focus on an instance which involves objects called *uniformisations*. A partial function $f \colon E \to F$ *uniformises* some relation $R \subseteq E \times F$ if it selects, for each $x \in E$, a particular y in F such that the pair $\langle x, y \rangle$ is in the relation R, whenever there exists one, see Figure 2.

It can be shown that the axiom of choice is true if and only if all relations admit uniformisations. This justifies the study of conditions under which uniformisations can be explicitly constructed for specific classes of relations.

Uniformisations are well-studied objects in *descriptive set theory*, the branch of mathematics that studies subsets of the real line or of other spaces that are, roughly, definable and well-behaved [Mos80, Kec95]. Many known results, stated in these two books, highlight the correspondences between the *constructive complexities* of the sets E and F, the relation R, and of an eventual uniformisation of it.

In this thesis, we decide to study this question of constructible uniformisations in the field of *formal languages*. The theory of formal languages is the domain that takes an interest in syntactic objects such as *words* or *trees*. These syntactic objects can be used in

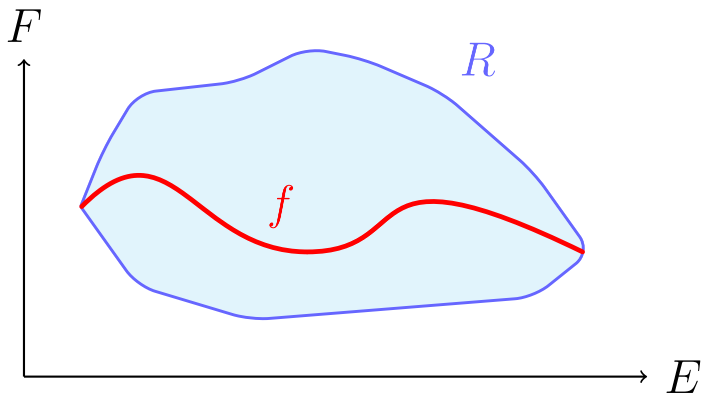
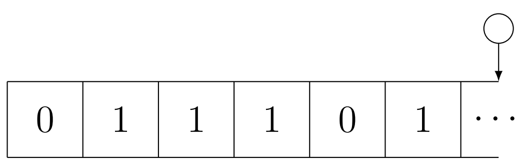


Figure 2: The function f uniformises the relation R.

numerous fields, including practical ones. In computer science, a word can represent a trace of some computation, while a tree can be used to describe different paths an algorithm can go. A *language*, *i.e.* a set of words or trees, is interpreted as a family of computations which we want to consider.



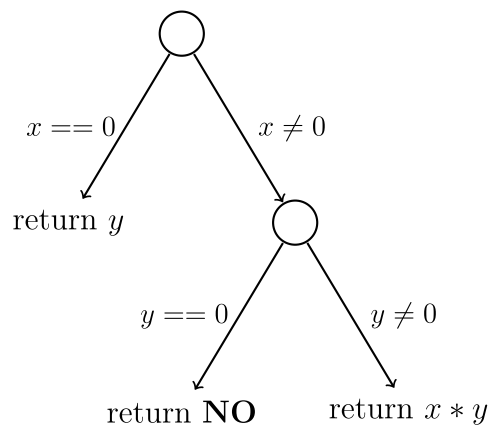


Figure 3: Words can represent outputs of a computation, while trees can represent the different possible paths of an algorithm.

To know which words, or trees, are contained in a language L, it is convenient to be able to define it by some property. For instance, L can be the language of computations that "output a 1 right after each 0". One of the most commonly used formalism to express such properties over words or trees is *Monadic Second-Order Logic* (MSO) [Büc60, Elg61, Tra62]. The reason why it is the most commonly used is that it is one of the most expressible yet still decidable formalism over words: there exists an algorithm that inputs an MSO sentence, and tells if there exists a word that satisfies it, and the same for trees [Rab69]. A language that can be defined via an MSO sentence is called *regular*.

As an example, here is how MSO over words can express Russel's example we gave above, about selecting infinitely many boots. In our vocabulary, we are given two letters: ℓ (for "left") and r (for "right"). An infinite word over this alphabet $\{\ell, r\}$ represents a selection of boots among the infinitely many pairs. For instance, the word $\ell \cdot r \cdot \ell \cdot \ell \cdot r \cdots$ represents the choice function that selects the left boot from the first pair, the right boot from the second pair, the left boot from the third pair, etc. A sentence like "At each position of the word, we have the letter r", which we write $\forall x.\ r(x)$ in MSO, defines exactly the infinite word $r \cdot r \cdot r \cdot r \cdots$, that represents the fact of selecting the right boot from each pair.

This leads to our main *regular uniformisation problem*. It asks if it is possible to define, in Monadic Second-Order Logic, a uniformisation of a given relation of words or trees.There already exists a wide and active field of research about regular uniformisations, in these contexts of words and trees.

For instance, we can cite [GS83], that presents this problem of regular uniformisations in a general way, and [LS98, RS08], that give an exhaustive survey about regular uniformisations in the fields of words and trees. A very important result is that, among finite words and infinite words, MSO has the ability to produce uniformisations for its own relations [LS98]:

Proposition 0.2. *Every regular relation of finite or infinite words admits a regular uniformisation.*

There are already known limitations of Proposition 0.2, since [GS83] and [CL07] prove that MSO does not admit this convenient *uniformisation property* when we consider the infinite binary tree:

Theorem 0.3 ([GS83])**.** *There exists a regular relation over the infinite binary tree that admits no regular uniformisation.*

Our main goal is to explore how Proposition 0.2 can be generalised.

Overview of the thesis

The thesis is composed as follows.

Chapter 1 is our preliminary chapter, in which are introduced all the notions we will be using all along the document. The concepts involved are: basic notions of set theory and of order theory, words and languages of words, First- and Monadic Second-Order Logics, algebra and its connections with languages of words, axiom of choice and uniformisations.

Then, in Chapter 2, we study different instances of the following problem:

Problem 0.4. *Given* $\mathbf{C}_1$ *and* $\mathbf{C}_2$ *two classes of regular languages, does every relation of* $\mathbf{C}_1$ *admit a uniformisation in* $\mathbf{C}_2$*?*

We show that, for many such classes included in **FO** (*i.e.* the class of languages definable in First-Order Logic), $\mathbf{C}_2$ is too weak to uniformise all relations of $\mathbf{C}_1$, and this holds even when $\mathbf{C}_2$ is more powerful than $\mathbf{C}_1$. From these results, which were originally published in [Mic18], follows the following proposition:

Proposition 0.5. *For any signature* Σ *contained in* $\{<, \mathsf{s}\}$*, neither* $\mathbf{FO}[\Sigma]$ *nor* $\mathbf{FO}^k[\Sigma]$*, for* $k \geq 1$*, admit the uniformisation property.*

The logics involved in the proposition are fragments of First-Order Logic: some can express the comparison between positions of words using the order $<$, some only have access to the relation s, which relates each pair of successive positions, while some only have access to letter tests. Finally, by $\mathbf{FO}^k$, we denote the logic where the number of distinct variables in the formulae is limited by the natural number k.

Although we know from the proposition above that the class $\mathbf{FO}[\,]$, meaning First-Order Logic with only letter tests, cannot uniformise all of its relations, we propose, in the same chapter, an algorithm testing if a given regular relation can be uniformised in it:

Proposition 0.6. *It is decidable whether a given regular language of finite words admits a uniformisation in* $\mathbf{FO}[\,]$*.*

This result opens the door to a whole class of problems:

Problem 0.7. *Let* $\mathbf{C}$ *be a class of regular languages. Is it decidable whether a given regular language admits a uniformisation in* $\mathbf{C}$*?*

From Proposition 0.5, together with Proposition 0.2, arises a question, whether there exists some robust subclass of regular languages of finite words, which can also uniformise its own relations. By *robust*, we mean a subclass which admits nice closure properties, such as closure under Boolean operations, closure under extensions of alphabets...

In Chapter 3, we answer negatively to this question when considering varieties of languages [Eil74], meaning subclasses of regular languages which are closed under Boolean operations, preimages under homomorphisms, and quotients:

Theorem 0.8. **MSO** *is the unique non-empty variety of languages that satisfies the uniformisation property.*

In other words, the ability of uniformising its own relations characterises the class **MSO** among all the non-empty robust classes. Therefore, this theorem, which was originally proven in [LMS19], gives another argument to the very particularity of the class of regular languages.

In Chapter 4, based on [MS20], we generalise Proposition 0.2 to *finitary linear orders*, which were first introduced in [LL66]. A linear order is *finitary* if it is obtained from singleton sets using a finite amount of the following operations:

- the operation $+$: $\lambda + \mu$ being the concatenation of the orders λ and μ;
- the operation $\times\omega$: $\lambda \times \omega$ consists of an infinite number of copies of λ concatenated one after the other $(\lambda + \lambda + \lambda + \cdots)$;
- the operation $\times\omega^\star$: $\lambda \times \omega^\star$ consists also of an infinite number of copies of λ, but in the other direction $(\cdots + \lambda + \lambda + \lambda)$;
- the operation η, which produces dense sets, such as $\mathbb{Q}$ (*i.e.* the set of *rational numbers*).

Finitary linear orders are of great interest because although they may be infinite, they admit a finite representation, and can be given as the input of an algorithm. Moreover, they are a representative way to approach countable linear orders, in the sense that any MSO sentence that is satisfied by a linear order is *a fortiori* satisfied by some finitary one [LL66].

We characterise the finitary linear orders λ admitting the regular-uniformisation property as the ones that do not admit any *non-trivial automorphisms*: a *non-trivial automorphism* of λ is a function from λ to itself which preserves the order and which is not the identity function. In fact, we prove that, considering a finitary linear order λ, the absence of such automorphisms is also equivalent to the possibility to regularly define natural objects such as *well orders* or choice functions, as states the following theorem, originally proven in [MS20]:

Theorem 0.9. *Let λ be a finitary linear order, then the following are equivalent:*

i) λ does not admit any non-trivial automorphism,

ii) every regular relation over λ admits a regular uniformisation,

iii) λ admits a regular choice function,

iv) λ admits a regular well order*,*

v) each element of λ is definable *in* MSO*.*

Moreover, Item i) is decidable, and, if it is true, then Items ii) to v) are constructible.

A *well order* over λ being an order $\sqsubseteq$ that does not admit any infinite decreasing sequence $x_0 \sqsupset x_1 \sqsupset x_2 \sqsupset \ldots$ Item $v)$ has to be understood as the following: for each element $x \in \lambda$, there exists a formula $\varphi_x^{\mathrm{def}}(y)$ such that for all $y \in \lambda$, $\varphi_x^{\mathrm{def}}(y)$ is satisfied if and only if $y = x$.

In a sense, Proposition 0.2 can now be seen as a particular case of Theorem 0.9, as finite and infinite words (which we call *ω-words* in the thesis) are finitary. As a consequence, the set $\mathbb{Z}$ of *integers* is an example of finitary linear order which does not satisfy the uniformisation property.

The crucial part of the theorem is the implication of Items *ii*) to *v*) by Item *i*), which we prove by introducing a notion of *condensation trees* for finitary linear orders, and navigating through them using algebraic tools introduced in [CCP18].

Finally, in Chapter 5, we study which implications of Theorem 0.9 hold and do not hold when removing the assumption of finitariness. As it turns out, the assumption of finitariness not only is important as a way of representing the given order but also plays an essential role in a number of implications of Theorem 0.9—without this assumption, the conditions are no longer equivalent.

Along the thesis, the proofs are mostly based on the algebraic approach to languages, via *semigroups* and *$\circ$-semigroups* (pronounced "circle"-semigroups). Hence, the whole thesis makes no mention of any *automaton*—another commonly used tool when it comes to regular languages [RS59]. However, most aspects of our algebraic approach can be equivalently seen from the automata-theoretic perspective.

Chapter 1

Preliminaries

In this first chapter, we introduce the main objects involved in the thesis. We also set the notations and the vocabulary that will be used throughout the whole document.

In Section 1.1, we recall some common notions of *set theory*, and introduce the mathematical objects which we work on in the whole thesis: *linear orders*, *words*, and *languages of words*. Then, in Section 1.2, we give a survey of *Monadic Second-Order Logic*, a formalism often used to express properties of words. In Section 1.3, we present some examples of *algebraic structures* and how they can be used to study languages of words. Finally, in Section 1.4, we expose the problem of *constructive choice*, and our objectives in regards to it.

1.1 Orders, words, and languages

1.1.1 Basic notions of set theory

Today, the context in which mathematics is most commonly studied is *Zermelo-Fraenkel set theory*, denoted by **ZF**, in which all mathematical objects are seen as *sets*. Although we do not go into a detailed description of the theory, we give in this subsection some basic notions of it. These notions are mostly standard, yet it is also the occasion to entirely fix their notations, as some may vary among the literature.

We call a *set* any collection E of mathematical objects. If x is an object of said collection, we write $x \in E$ and say that x is an *element* of E, that x *belongs* to E, that x *is* in E, or that E *contains* x. Sets are defined by their elements and are said *equal* (written $E = F$) if they contain exactly the same elements, and we generally write them with their elements between braces: $E = \{x, y, \dots\}$. There exists a unique set, the *empty set*, denoted by $\emptyset$, that does not contain any element (*i.e.* $\emptyset = \{\}$), and all sets that contain at least one element are called *non-empty*. A set that contains a unique element is called a *singleton*.

An important set is the set $\mathbb{N} := \{0, 1, 2, \dots\}$ of *natural numbers*, usually with its *addition*, written $+$, and its *multiplication*, written $\times$. We say that a natural number is *positive* if it is not 0.

Let E and F be two sets. Then $E \cup F$ denotes the *union* of E and F, *i.e.* the set of elements that belong to E or to F (not only here but in the whole thesis, this "or" has to be understood in the inclusive sense: said elements can belong to both E and F), and $E \cap F$ denotes the *intersection* of E and F, *i.e.* the set of elements that belong to both E and F. The sets E and F are said *disjoint* if $E \cap F = \emptyset$, meaning if no mathematical object is an element of both E and F. Under this assumption, we rather write $E \sqcup F$ for $E \cup F$, to emphasise that the union is disjoint. These notions of union and intersection are extended to more than two sets: if I is a set, and for each $i \in I$, so is E_i, then $\bigcup_{i \in I} E_i$ (resp. $\bigcap_{i \in I} E_i$) denotes the set of elements that belong to at least one of the E_i's (resp. to all of the E_i's). Similarly, if the E_i's are assumed pairwise disjoint ($E_i \cap E_j = \emptyset$ for all $i \neq j$), then we rather write $\bigsqcup_{i \in I} E_i$ for $\bigcup_{i \in I} E_i$. Finally, we denote by $E \setminus F$ the set of elements that are in E and not in F.

We write $E \subseteq F$ to say that the set E is *included* in the set F, meaning that each element of E is also an element of F. It is equivalent to saying that the set $E \setminus F$ is empty. In this case we also say that E is a *subset* of F. The set of subsets of E is written $\mathcal{P}(E)$. We denote by $E \nsubseteq F$ the fact that E is not included in F, meaning that there exists some element of E that is not an element of F. We also use the symbol $\subset$ for the *strict inclusion*: we write $E \subset F$ if $E \subseteq F$ and moreover $E \neq F$, meaning that there exists some element x of F that is not an element of E.

We denote by $E \times F$ the *product set* of E and F, meaning the set of *pairs* $\langle x, y \rangle$ with x being an element of E and y an element of F. We write E^2 for the product set $E \times E$, and, more generally, if n is a positive natural number, we write E^n for the product $E \times E \times \cdots \times E$, with E appearing exactly n times. It is the set of *n-tuples* over E, meaning the set of elements of the shape $\langle x_0, x_1 \dots, x_{n-1} \rangle$, where each x_i is an element of E.

A *binary relation* R between E and F is a subset of $E \times F$. In this case, we more often write xRy rather than $\langle x, y \rangle \in R$. If $E = F$, we say that R is a binary relation over E. In the same way, a subset of E^n, for some positive natural number n, is called an *n-ary relation* over E. If E' is a subset of E, then we denote by $R\restriction_{E'}$ the *restriction* of R to E': it is the binary relation between E' and F such that for all $\langle x, y \rangle \in E' \times F$, we have $xR\restriction_{E'}y$ if and only if xRy.

A *function* from E to F is a binary relation f between E and F such that for all $x \in E$, there exists a unique $y \in F$ such that xfy. We denote by $f(x)$ this said y, and say that f

maps x to y. This y is also called the *image* of x under f, and x is called a *preimage* of y under f. We write $f\colon E \to F$ to tell that f is a function from E to F, and $f\colon x \mapsto y$ (or sometimes $x \overset{f}{\mapsto} y$) to express that x is mapped to y by f. Notice that the restriction $f{\restriction_{E'}}$ of f to a subset E' of E is also a function, from E' to F. Two functions f and g from E to F are said *equal* (denoted by $f = g$) if they map the same elements of E to the same elements of F, *i.e.* $f(x) = g(x)$ for every $x \in E$. For each subset E' of E, we denote by $f(E')$ the set $\{f(x) \mid x \in E'\}$ of images of elements of E', and for each subset F' of F, we denote by $f^{-1}(F')$ the set $\{x \in E \mid f(x) \in F'\}$ of preimages of elements of F'. For each set E, there exists a particular function called the *identity* of E. It is the function mapping each element x of E to itself. We denote it by id_E. If f is a function from E to F, and g a function from F to a set G, then by $g \circ f$ we denote the *composition* of g and f, it is the function from E to G that maps every $x \in E$ to the element $g(f(x)) \in G$. For every function f from E to F, we have the equalities $f \circ \mathrm{id}_E = f$, and $\mathrm{id}_F \circ f = f$.

Let f be a function from E to F. If every element of F admits at most one preimage under f, then f is called *injective* (or an *injection*), and if every element of F admits at least one preimage under f, then f is called *surjective* (or a *surjection*). It is easy to see that the composition of two injective functions (resp. of two surjective functions) is also injective (resp. surjective). If f is both injective and surjective, meaning that each element of F admits exactly one preimage under f, then it is called *bijective* (or a *bijection*). In this case, the function from F to E that maps each $y \in F$ to the unique $x \in E$ such that $f(x) = y$ is a bijection from F to E. This function, denoted by f^{-1}, is called the *inverse* function of f, and has the property that $f^{-1} \circ f = \mathrm{id}_E$ and $f \circ f^{-1} = \mathrm{id}_F$. Since the composition of two injections is an injection and the composition of two surjection is a surjection, we know that the composition of two bijection is also a bijection.

A function f from E to F can also be denoted by $(f(x))_{x\in E}$, in which case it is called a *family*. Finally, a *partial function* from E to F is a function f from a subset of E to F. In this case, said subset is called the *domain* of f, and is denoted by $\mathrm{Dom}(f)$.

A set E is *finite* if there exists a bijection from E to the set $\{0, 1, \ldots, n-1\}$ for some natural number $n \in \mathbb{N}$. Said set $\{0, 1, \ldots, n-1\}$ is usually identified with n itself, which is called the *cardinality* of E, and is denoted by $|E|$. Finally, E is *countable* if either it is finite or there exists a bijection from E to the set of natural numbers $\mathbb{N}$.

In the thesis, we will often meet the term *class*. A *class* is also a collection of mathematical objects, simply less formal. We do not state a rigorous definition of classes, in comparison to sets. It is enough for the reader to remember that the notions we defined in this subsection do not always make sense when considering classes.

1.1.2 Equivalence relations and orders

Let R be a binary relation over a set E. It is *reflexive* if for every $x \in E$, we have xRx. On the contrary, it is *antireflexive* if xRx is true for no element $x \in E$. It is *transitive* if for all $x, y, z \in E$, if xRy and yRz then xRz. It is *symmetric* if for all $x, y \in E$, xRy implies yRx. Finally, it is *antisymmetric* if for all $x, y \in E$, if xRy and yRx then $x = y$.

A binary relation over E that is reflexive, transitive, and symmetric is called an *equivalence relation* over E. In this thesis, we will denote equivalence relations by symbols such as $\equiv$, or $\cong$. If $\equiv$ is an equivalence relation over a set E, we say that two elements x and y of E are $\equiv$-*equivalent*, or simply *equivalent*, if $x \equiv y$. If $x \in E$, then we define $[x]_\equiv$ as the set $\{y \in E \mid x \equiv y\}$, the *equivalence class* of x. We denote by $E/\equiv$ the set $\big\{[x]_\equiv \mid x \in E\big\}$ of all these equivalence classes.

A binary relation over E that is transitive, reflexive (resp. antireflexive), and antisymmetric is called an *order* (resp. a *strict order*) over E. Orders (resp. strict orders) are generally denoted by the symbol $\leq$ (resp. $<$). Any order $\leq$ over E induces a strict order $<$, defined by $x < y$ if $x \leq y$ and moreover $x \neq y$. Reciprocally, any strict order $<$ over E induces an order $\leq$, defined by $x \leq y$ if either $x < y$ or $x = y$. Therefore, we often identify the two notions, the symbols $\leq$ and $<$ indicating whether we consider the pairs $\langle x, x\rangle$ in the relation or not. In this thesis, an ordered set, meaning a set with an order, is generally denoted by the letter ϑ and identified with its order. If ϑ is an ordered set, and $x, y \in \vartheta$, then we write $x \geq y$ for $y \leq x$ and $x > y$ for $y < x$. If X is a subset of ϑ, then it can be provided with the order of ϑ restricted to X^2, meaning $\leq\restriction_X := \big\{\langle x, y\rangle \mid x, y \in X \text{ and } x \leq y\big\}$. It is clear that this new relation is an order over X, which explains why a subset of an order is sometimes called a *suborder*. We generally simply write it $\leq$ rather than $\leq\restriction_X$.

Let ϑ be an ordered set. A *maximal* (resp. *minimal*) *element* of ϑ is an element $m \in \vartheta$ such that for all $x \in \vartheta$, if $m \leq x$ (resp. $x \leq m$) then $x = m$. A *greatest element* (resp. *least element*) of ϑ is an element $m \in \vartheta$ such that $x \leq m$ (resp. $m \leq x$) for all $x \in \vartheta$. Such an element of ϑ, if it exists, is necessarily unique. Notice that a greatest (resp. least) element is necessarily maximal (resp. minimal), but the contrary is not true in general. The notion of maximal elements naturally extends to any suborder X of ϑ: a maximal element of X is an element $m \in X$ such that for all $x \in X$, if $m \leq x$ then $x = m$, and, the same way, we extend to X the notions of minimal, greatest, or least elements.

Let ϑ be an order. We say that it is *linear*, or *total*, if all elements are *comparable* with each other, meaning that for all $x, y \in \vartheta$, we have either $x \leq y$ or $x \geq y$. Elements of a linear order are called *positions* of this order. In this thesis, we mostly denote a linearly ordered set by the letter λ. Naturally, any suborder of a linear order is linear as well. Notice that in a linear order λ, the notions of maximal and greatest elements coincide, the same for the notions of minimal and least elements. A linear order λ is *dense* if it has at least two distinct elements and if for all $x < z$ in λ, there exists some $y \in \lambda$ with $x < y < z$. If λ does not admit any dense subset, it is called *scattered*. A subset X of λ is *convex* if for all $x \leq y \leq z \in \lambda$, if x and z are in X then so is y. Finally, if X and Y are two subsets of λ, we denote by $X < Y$ the fact that $x < y$ for all elements $x \in X$ and $y \in Y$.

If an order has the property that each of its non-empty suborders admits a least element, then it is called a *well order*. Well orders are generally denoted by the letter ϖ in this thesis. Naturally, any suborder of a well order is also a well order. In particular, any well order ϖ is necessarily linear, because each subset $\{x, y\}$ must admit a least element. A family defined on a linear order is called a *sequence*.

Let ϑ_1 and ϑ_2 be two orders. An *isomorphism* from an ϑ_1 to ϑ_2 is a bijection ι from ϑ_1 to ϑ_2 such that for all $x, y \in \vartheta_1$, we have $x < y$ if and only if $\iota(x) < \iota(y)$. We say that it *preserves* the order. If there exists such an isomorphism, these orders are said *isomorphic* (to each other). Isomorphisms preserve the properties of being linearly and well ordered: if ϑ_1 and ϑ_2 are two isomorphic orders, and if ϑ_1 is a linear order (resp. a well order), then so is ϑ_2. In the thesis, we will sometimes identify isomorphic orders with each other, and whenever that happens, we will explicitly write it. An isomorphism from an order ϑ to itself is called an *automorphism* of ϑ. A particular automorphism of ϑ is id_ϑ, the identity function of ϑ. We say that an automorphism of ϑ is *non-trivial* if it is not this identity function id_ϑ. Finally, we notice that the composition of two isomorphisms, and the inverse of an isomorphism are also isomorphisms.

We say that an order is *finite* (resp. *countable*) if it is as a set. A few linear orders are given a particular attention: the set $\mathbb{N}$ of natural numbers itself, which we will denote by ω to emphasise that we provide it with its usual order ($0 < 1 < 2 < \dots$); the set $\omega^\star := \{\dots, -3, -2, -1\}$ of *negative integers*; the set $\mathbb{Z} := \omega^\star \sqcup \omega$ of *integers*; and $\mathbb{Q}$, the set of *rational numbers*. All of them are with their usual orders. Notice that ω is a well order, but the other three are not.

We state two theorems linking $\mathbb{Q}$ with all the other countable linear orders:

Theorem 1.1 (Cantor)**.** *Any countable linear order is isomorphic to a suborder of* $\mathbb{Q}$*.*

A proof of Theorem 1.1 can be found in [Ros81, Theorem 2.5]. It is a well-known result that, under the additional assumption that the countable linear order is also dense and without least nor greatest element, the same proof can be modified to obtain an isomorphism from the linear order to the whole set $\mathbb{Q}$:

Theorem 1.2. *Up to isomorphism, $\mathbb{Q}$ is the only countable linear order that is dense and without least nor greatest element.*

1.1.3 Words

An *alphabet* is any non-empty set $\mathbb{A}$, and we call its elements *letters*. The alphabets we consider in the thesis are usually finite, unless explicitly stated otherwise. Let $\mathbb{A}$ be an alphabet and λ be a linear order. Any function w from λ to $\mathbb{A}$ is called a λ-*word* over $\mathbb{A}$, or more simply a *word* over $\mathbb{A}$. A *position* of w is simply a position of λ, and it is said to be *labelled* by a letter $a \in \mathbb{A}$ (in w) if $w(x) = a$. We call λ the *domain* of w, and we denote it by $\mathrm{Dom}(w)$. The unique word of empty domain $\emptyset$ is written ϵ. If $\mathbb{A}$ is a singleton, then we identify w with its domain $\mathrm{Dom}(w)$. If X is a subset of $\mathrm{Dom}(w)$, then we denote by $w\restriction_X$ the word over $\mathbb{A}$ of domain X defined by $w\restriction_X(x) = w(x)$ for all $x \in X$. If moreover X is convex, $w\restriction_X$ is called a *factor* of w.

Let w_1 and w_2 be two words over an alphabet $\mathbb{A}$. An *isomorphism* from w_1 to w_2 is an isomorphism ι from $\mathrm{Dom}(w_1)$ to $\mathrm{Dom}(w_2)$ that additionally preserves the labels, meaning that for all $x \in \mathrm{Dom}(w_1)$, $w_2(\iota(x)) = w_1(x)$. Two words over $\mathbb{A}$ are called *isomorphic* (to each other) if there exists an isomorphism from one to the other. Similarly as for orders, we will sometimes identify isomorphic words with each other, and whenever we do so, it will be explicitly noted. An isomorphism from a word w to itself is called an *automorphism*, and we call it *non-trivial* if is is not the identity function $x \mapsto x$. The same way, the composition of two isomorphisms, and the inverse of an isomorphism, are also isomorphisms.

If λ is a linear order, the set of λ-words over $\mathbb{A}$ is denoted $\mathbb{A}^\lambda$, and a λ-*language* over $\mathbb{A}$ is any subset of $\mathbb{A}^\lambda$. If L is a λ-language over $\mathbb{A}$, then we define L^{c}, the *complement* of L, as the λ-language $\mathbb{A}^\lambda \backslash L$, this operation, together with the union $\cup$ and the intersection $\cap$, forms the *Boolean operations* of $\mathbb{A}^\lambda$. When the linear order λ is clear in the context, a λ-language is simply called a *language*.

In this thesis, we simply denote a λ-language by the letter L, *i.e.* without any explicit mention of the alphabet. In most cases, this will not be a problem since the considered alphabet will be clear. Yet, in some cases, in particular in Subsection 1.2.4 and in the first section of Chapter 3, we will want to avoid any possible ambiguity, and we will write $\langle L, \mathbb{A}\rangle$ for L, to make clear that it shall be considered as a language over $\mathbb{A}$ and not over any bigger alphabet.

Finite words

Words of finite domains are of particular interest in this thesis. We call them *finite words*. Up to isomorphism, we always suppose that the domain of a finite word w is the set $\{0, \dots, n-1\}$ itself, for a particular $n \in \omega$. The number n in question is also called the *length* of w, and denoted by $|w|$. For each letter $a \in \mathbb{A}$, we also denote by $|w|_a$ the cardinality of the set $\{x \in \mathrm{Dom}(w) \mid w(x) = a\}$, meaning the number of occurrences of a in w.

In this thesis, for the sake of convenience, we make the choice to not take the empty word into consideration when finite words are involved, unless explicitly stated otherwise. With this convention, the set of finite words is denoted by $\mathbb{A}^+$, *i.e.* $\mathbb{A}^+ := \bigsqcup_{n\in\mathbb{N}\setminus\{0\}} \mathbb{A}^n$, and by $\mathbb{A}^*$ we denote the set of possibly empty finite words over $\mathbb{A}$, *i.e.* $\mathbb{A}^* := \bigsqcup_{n\in\mathbb{N}} \mathbb{A}^n = \mathbb{A}^+ \sqcup \{\epsilon\}$. We call a *language* of finite words over $\mathbb{A}$ any subset of $\mathbb{A}^+$ (or, once again, of $\mathbb{A}^*$ when it is explicitly stated than the empty word is considered). Thus, we also call $\emptyset$ the *empty language* and $\mathbb{A}^+$ the *full language* over the alphabet $\mathbb{A}$. Although alphabets are non-empty, there will be along this thesis some constructions involving the notations $\emptyset^+$ and $\emptyset^*$: the former is understood as the empty language, while the latter is understood as $\{\epsilon\}$, the language composed of a unique word, the empty word. The *Boolean operations* of $\mathbb{A}^+$ are composed of the union $\cup$, the intersection $\cap$, and the complement $L \mapsto L^{\mathrm{c}}$, where here L^{c}, the *complement* of L, denotes here the language $\mathbb{A}^+ \setminus L$.

If w_1 and w_2 are two (possibly empty) finite words, over $\mathbb{A}$ and $\mathbb{B}$ respectively, of lengths p and q respectively, then we define the *concatenation* (or *product*) of w_1 and w_2, written $w_1 \cdot w_2$, as the finite word w over $\mathbb{A}\cup\mathbb{B}$, of length $p+q$, defined by $w(i) = w_1(i)$ for $0 \leq i < p$ and $w(i) = w_2(i-p)$ for $p \leq i < p+q$. We often use this notation to write $w(0)\cdot w(1)\cdots w(|w|-1)$ for a finite word w, identifying any word of length 1 labelled by the letter a with the letter a itself. If w is a finite word, and $n \in \mathbb{N}$, then we denote by w^n the product $w \cdot w \cdots w$, where w appears exactly n times (if $n = 0$ then it is the empty word ϵ). We call it the *exponentiation* of w to the power n.

These notions of product and exponentiation are extended to languages: if L_1 and L_2 are two languages of (possibly empty) words, over $\mathbb{A}$ and $\mathbb{B}$ respectively, then $L_1 \cdot L_2$ denotes the *concatenation* (or *product*) of L_1 and L_2, defined as the language of finite words w over $\mathbb{A}\cup\mathbb{B}$ that can be written as $w_1 \cdot w_2$, with $w_1 \in L_1$ and $w_2 \in L_2$. The same way, if L is a language of finite words over $\mathbb{A}$ and $n \in \mathbb{N}$, then L^n denotes the *exponentiation* of L to the power n, defined as the set of words over $\mathbb{A}$ that can be written as $w_0 \cdot w_1 \cdots w_{n-1}$, with each $w_i \in L$.

If the alphabet $\mathbb{A}$ is linearly ordered, then we can also define a linear order on $\mathbb{A}^*$. We call it the *lexicographic order*, written $\leq_{\mathrm{lex}}$. It is defined by $w_1 \leq_{\mathrm{lex}} w_2$ if either w_1 is a *prefix* of w_2 (meaning that $w_2 = w_1 \cdot w_3$ for some word $w_3 \in \mathbb{A}^*$), or there exists some position $k \in \mathrm{Dom}(w_1) \cap \mathrm{Dom}(w_2)$ such that $w_1(k) < w_2(k)$ and $w_1(i) = w_2(i)$ for all $i < k$. In particular, $\epsilon \leq_{\mathrm{lex}} w$ for all words $w \in \mathbb{A}^*$. If $\mathbb{A}$ is finite (and, once again, this is the main assumption in the thesis, even when we do not specify it explicitly), then this order restricted to each set $\mathbb{A}^n$, for $n \in \mathbb{N}$, is a well order. This is a clear consequence of the fact that $\mathbb{A}^n$ is finite. However, except in the trivial case where $\mathbb{A}$ is a singleton, this statement becomes false for the full set $\mathbb{A}^*$, for if a and b are two distinct letters in $\mathbb{A}$ ordered by $a < b$, then the language $\{a^i \cdot b \mid i \in \mathbb{N}\} \subseteq \mathbb{A}^*$ does not admit a lexicographically least word (because $a^{i+1} \cdot b <_{\mathrm{lex}} a^i \cdot b$ for every $i \in \mathbb{N}$).

ω-words and $\mathbb{Q}$-words

Another important class of words in this thesis is the class of *ω-words.* An *ω-word* is a word whose domain is isomorphic to ω, the ordered set of natural numbers. Once again, up to isomorphism, we always suppose that the domain of any ω-word is ω itself. We extend the notions of concatenations and exponentiations defined in the previous paragraph: if w_1 is a finite word over $\mathbb{A}$ and w_2 is an ω-word over $\mathbb{B}$, then $w_1 \cdot w_2$, the *concatenation* of w_1 and w_2, and w_1^ω, the *exponentiation* of w_1 to the power ω, are also two ω-words. Formally, if p is the length of w_1, the former is defined as the ω-word w over $\mathbb{A} \cup \mathbb{B}$ with each position $i \in \omega$ labelled by $w_1(i)$ if $0 \leq i < p$ and by $w_2(i - p)$ if $p \leq i$, while the latter is defined as the ω-word w over $\mathbb{A}$ with each position $i \in \omega$ labelled by $w_1(i[p])$, $i[p]$ being the unique $j \in p$ such that $i = p \times k + j$ for some $k \in \mathbb{N}$. Notice that we can allow w_1 to be the empty word in the definition of $w_1 \cdot w_2$, but not in the definition of w_1^ω.

We can also define a *lexicographic order* $\leq_{\mathrm{lex}}$ on the set of ω-words over a linearly ordered alphabet $\mathbb{A}$. It is defined by $w_1 \leq_{\mathrm{lex}} w_2$ if either $w_1 = w_2$ or there exists a position $k \in \omega$ such that $w_1(k) < w_2(k)$ and $w_1(i) = w_2(i)$ for all $i < k$. Similarly as for $\mathbb{A}^+$, $\leq_{\mathrm{lex}}$ is a linear order but not a well order (unless $\mathbb{A}$ is a singleton).

In this thesis, we will also consider $\mathbb{Q}$-words, meaning words whose domain is (isomorphic to) the set of rational numbers. The proof of Theorem 1.2 can be generalised, in order to show that for any alphabet $\mathbb{A}$, there exists, up to isomorphism, a unique $\mathbb{Q}$-word w over $\mathbb{A}$ that is *densely labelled*, min the sense that for each positions $p < r$ and each letter $a \in \mathbb{A}$, there exists some position q labelled by a in w with $p < q < r$. We call this $\mathbb{Q}$-word the *perfect shuffle* over $\mathbb{A}$, and denote it by $\mathbb{A}^\eta$.

Countable and finitary words

A word over an alphabet $\mathbb{A}$ whose domain is countable is called a *countable word*. The class of countable words over $\mathbb{A}$ is denoted by $\mathbb{A}^\circ$. As for finite words, when we introduce a countable word, we consider it non-empty by default, unless we state it otherwise. As Theorem 1.1 states that any countable linear order is isomorphic to some subset of $\mathbb{Q}$, set theory allows us to treat $\mathbb{A}^\circ$ as a set, and we call a *language* of countable words any subset of $\mathbb{A}^\circ$. As usual, we provide $\mathbb{A}^\circ$ with *Boolean operations*: $\cup$, $\cap$, and $L \mapsto L^{\mathrm{c}}$, which here denotes the set $\mathbb{A}^\circ \setminus L$.

In this paragraph, we generalise the notion of concatenation which we previously defined for finite and ω-words. Let I be a set, which we suppose linearly ordered, and, for each $i \in I$, let w_i be a countable word over $\mathbb{A}$. We define the *generalised concatenation*, or simply *concatenation*, of the w_i's as being the word w over $\mathbb{A}$ of domain $\bigsqcup_{i \in I} \big\{ \langle x_i, i \rangle \mid x_i \in \mathrm{Dom}(w_i) \big\}$, such that $w(\langle x_i, i \rangle) = w_i(x_i)$ for each $i \in I$ and $x_i \in \mathrm{Dom}(w_i)$. The domain is linearly ordered by $\langle x_i, i \rangle \leq \langle y_j, j \rangle$ if either $i < j$, or $i = j$ and $x_i \leq y_i$ in $\mathrm{Dom}(w_i)$. We denote by $\sum_{i \in I} w_i$ this concatenation. Since words over a singleton alphabet are identified with their domains, we deduce a definition of a *generalised concatenation* of linear orders.

We use special notations in some particular cases. If I is (isomorphic to) $\{1, 2\}$, then this concatenation is written $w_1 \cdot w_2$. Note that if w_1 is a finite word and w_2 is either a finite word or an ω-word, it boils down to our previous definitions of concatenation. If I is ω (resp. $\omega^\star$), and if all the w_i's are isomorphic to a same word w, then this concatenation is written w^ω (resp. $w^{\omega^\star}$), and we call it the *exponentiation* of w to the power ω (resp. to the power $\omega^\star$). Once again, it coincides to our previous definition of exponentiation if w is a finite word. Similarly, we write w^n in the case $I = n$ and all the w_i's are isomorphic to w. Finally, if $w_0, \ldots, w_{n-1}$ are words over $\mathbb{A}$, then $\{w_0, \ldots, w_{i-1}\}^\eta$ denotes the word $\sum_{q \in \mathbb{Q}} w_{u(q)}$, where $u = \{0, \ldots, n-1\}^\eta$ is the perfect shuffle of $\{0, \ldots, n-1\}$. This word is called the *perfect shuffle* of the w_i's.

These concatenations and exponentiations allow us to give finite representations to some countable words: *finitary words*. These words, first introduced in [LL66], are of great interest, since they can be put in algorithms and can efficiently approach unrestricted countable words (see the same article).

A *finitary word* over $\mathbb{A}$ is a word that can be constructed from single letters using finitely many times the operations $\cdot$, $(\cdot)^\omega$, $(\cdot)^{\omega^\star}$, and $(\cdot)^\eta$. More formally: for each $a \in \mathbb{A}$, the word composed of a single position labelled by a is finitary; if w_1 and w_2 are both finitary, then so is $w_1 \cdot w_2$; if w is a finitary, then so are w^ω and $w^{\omega^\star}$; if $w_0, \ldots, w_{n-1}$ are all finitary, with $n \in \mathbb{N}$, then so is $\{w_0, \dots, w_{n-1}\}^\eta$. In particular, a finitary word is countable. As an example, the $\mathbb{Z}$-word over $\{a, b\}$ in which each even position is labelled by the letter a and each odd position is labelled by the letter b is finitary, because it is obtained as $(a \cdot b)^{\omega^\star} \cdot (a \cdot b)^\omega$. On the contrary, the ω-word $\Sigma_{i\in\omega} a \cdot b^i$, consisting of i letters b between the i-th and the $(i+1)$-th a is an example of a word that is not finitary, since it is does not fulfil the condition stated in the following fact:

Fact 1.3. *A finitary ω-word over $\mathbb{A}$ is necessary of the shape $u \cdot v^\omega$, with $u \in \mathbb{A}^*$ and $v \in \mathbb{A}^+$.*

The proof of this fact involves *cuts*: a *cut* of a linear order λ is a subset $X \subseteq \lambda$ that is *closed downward*, meaning that for all $x \leq y \in \lambda$, if $y \in X$ then $x \in X$. It is *strict* if it is not λ itself.

Proof. Let us first notice that each strict cut of ω is finite. Therefore, the operations $(\cdot)^{\omega^\star}$ and $(\cdot)^\eta$ cannot be used to obtain an ω-word, since they would involve infinite strict cuts. Similarly, one cannot use the operation $(\cdot)^\omega$ more than once, and the fact follows. □

Since a word over a singleton alphabet is identified with its domain, the above operations can be used to construct linear orders, and we deduce a notion of *finitary linear orders*. In this case, we rather write $\lambda_1 + \lambda_2$ for $\lambda_1 \cdot \lambda_2$ and $\lambda \times \omega$ (resp. $\lambda \times \omega^\star$) for λ^ω (resp. $\lambda^{\omega^\star}$), to keep the operations consistent with the ordinal-theoretic ones (see [Sie58]). The previous example $\Sigma_{i\in\omega} a \cdot b^i$ shows that a non-finitary word can yet have a finitary domain. Like for words, a finitary linear order is necessarily countable, and, finally, a linear order obtained only via the operations $+$, $\times\omega$, and $\times\omega^\star$ is necessarily scattered: only the operation $(\cdot)^\eta$ can create dense subsets.

Fact 1.4. *A finitary linear order whose construction does not involve the operation $(\cdot)^\eta$ is scattered.*

Indeed, one clearly sees that the concatenation preserves the scattered property. The same for the two other operations $\times\omega$ and $\times\omega^\star$. Moreover, Haussdorf's theorem [Hau08] characterises the countable linear orders that are scattered as the ones obtained from singleton sets and using generalised concatenation over the linear orders $\{1, 2\}$, ω, and $\omega^\star$.

1.2 Logic and regularity

Most of our objectives concern the ability to express some properties on linear orders and on languages of words in certain formalisms, which we present in this subsection. In Subsections 1.2.1 and 1.2.2, we give a formal definition of *Modal Second-Order Logic* and some of its common fragments, such as *First-Order Logic*. Then, in Subsection 1.2.3, we introduce *Ehrenfeucht-Fraïssé games*, important tools to disprove the possibility to express some properties in said fragments. Finally, in Subsection 1.2.4, we see how Monadic Second-Order Logic and its fragments are understood in the framework of words, and we define at the same time the classes of languages which we study in the thesis.

1.2.1 Model theory and Monadic Second-Order Logic

The formalisms which we work on are characterised by a *grammar*, *i.e.* a family of rules allowing the construction of mathematical sentences, and by a *signature*, that consists of the mathematical objects allowed as a "basis" of the sentences. For the sake of completeness, we give here a complete survey of these notions, and also of the link between the sentences in the formalism—the *syntax*, and their interpretations in an actual mathematical model—the *semantics*. These definitions are standard and can be found for instance in [Sha91].

A *signature* Σ is a set of symbols, each given with its own *arity* $n \in \mathbb{N}$, meaning its own number of arguments. Some are called *functional symbols* (generally written with lower case letters such as $c, f, g, h \dots$), and others are called *relational symbols* (generally written with capital letters such as $P, Q, R, S \dots$). A functional symbol of arity 0 is called a *constant symbol*, and a relational symbol of arity 1 is called a *predicate symbol*. A *model* of Σ is a set $\mathcal{M}$ with an *interpretation* of each symbol $s \in \Sigma$, written $s^{\mathcal{M}}$ and given in the following way: the interpretation of a functional symbol f of arity n is a function $f^{\mathcal{M}} \colon \mathcal{M}^n \to \mathcal{M}$, and the interpretation of a relational symbol R of arity n is an n-ary relation $R^{\mathcal{M}} \subseteq \mathcal{M}^n$. In particular, the interpretation of a constant symbol c is an element $c^{\mathcal{M}} \in \mathcal{M}$. It is generally clear whether we speak about a symbol in the signature or about its interpretation in a model, therefore, we will often simply write s for $s^{\mathcal{M}}$. By default, all the signatures involved in the thesis contain the *equality symbol* $=$, even if we do not explicitly state it. The interpretation $=^{\mathcal{M}}$ of this symbol is the usual relation $\big\{\langle x, x\rangle \mid x \in \mathcal{M}\big\}$.

Let $\mathcal{M}$ and $\mathcal{N}$ be two models of a signature Σ. A function ι from $\mathcal{M}$ to $\mathcal{N}$ is an *isomorphism* if it is bijective and if it preserves the interpretations of the symbols of Σ, *i.e.* $\iota\big(f^{\mathcal{M}}(m_0, \dots, m_{n-1})\big) = f^{\mathcal{N}}\big(\iota(m_0), \dots, \iota(m_{n-1})\big)$ for each functional symbol f of Σ of arity n, and all $m_0, \dots, m_{n-1}$ in $\mathcal{M}$, and $R^{\mathcal{M}}(m_0, \dots, m_{n-1})$ if and only if $R^{\mathcal{N}}\big(\iota(m_0), \dots, \iota(m_{n-1})\big)$ for each relational symbol R of arity n and all $m_0, \dots, m_{n-1}$ in $\mathcal{M}$.

We are given a countable number of symbols, which we call the *first-order variables*: $x, y, z, \dots$, and another, disjoint, whose symbols are called *monadic second-order variables*, or simply *second-order variables*: $X, Y, Z, \dots$ Given a signature Σ, we define inductively a notion of *terms* over Σ (or simply *terms*): each first-order variable x is a term; if $c \in \Sigma$ is a constant symbol, then c is a term; and if $t_0, \ldots, t_{n-1}$ are terms and $f \in \Sigma$ is a functional symbol of arity n, then also $f(t_0, \ldots, t_{n-1})$ is a term.

Now, with this notion of terms, we define inductively the notion of *Monadic Second-Order formulae* over Σ, whose class forms the *Monadic Second-Order Logic* over Σ, written MSO[Σ] (or simply MSO): if $t_0, \ldots, t_{n-1}$ are terms and $R \in \Sigma$ is a relational symbol of arity n, then $R(t_0, \ldots, t_{n-1})$ is a formula; if t is a term and X a second-order variable, then $t \in X$ is a formula; if φ is a formula, then so is $\neg\varphi$; if φ_1 and φ_2 are two formulae, then so is $\varphi_1 \vee \varphi_2$; if φ is a formula and x is a first-order variable, then $\exists x.\ \varphi$ is a formula; finally, if φ is a formula and X is a second-order variable, then $\exists X.\ \varphi$ is a formula. A formula of the shape $R(t_0, \ldots, t_{n-1})$ is called *atomic*, the formula $\neg\varphi$ is called the *negation* of φ, the formula $\varphi_1 \vee \varphi_2$ is the *disjunction* of φ_1 and φ_2, and the symbol $\exists$ is called the *existential quantifier* (and is of the same order as the variable that follows it).

The precedence of the constructors is the following: $\neg$ binds stronger than $\vee$, which binds stronger than the existential quantifier. Hence, for instance, if P is a predicate symbol and φ a formula, then $\exists x.\ \neg P(x) \vee \varphi$ shall be understood as the formula $\exists x.\ \big((\neg P(x)) \vee \varphi\big)$. Also, we consider the precedence to the left for different constructors $\vee$: if φ_1, φ_2, and φ_3 are three formulae, then $\varphi_1 \vee \varphi_2 \vee \varphi_3$ shall be understood as the formula $(\varphi_1 \vee \varphi_2) \vee \varphi_3$. Finally, if $(\varphi_i)_{i\in n}$ is a family of formulae, we write $\bigvee_{i\in n} \varphi_i$ for $\varphi_0 \vee \varphi_1 \vee \cdots \vee \varphi_{n-1}$.

To each term t over Σ, we assign the set $Var(t)$ of variables that occur in t. It is formally defined inductively, by: $Var(c) = \emptyset$ for each constant symbol c, $Var(x) = \{x\}$ for each first-order variable x, and $Var(f(t_0, \ldots, t_{n-1})) = \bigcup_{i\in n} Var(t_i)$ for every n-ary functional symbol f and all terms $t_0, \ldots, t_{n-1}$.

Now, we assign, to each MSO[Σ] formula φ, the set $FreeVar(\varphi)$ of its *free variables*: it is the set of variables (both first- and second-order) that occur in φ but that are not introduced by an existential quantifier $\exists$. More formally, it is inductively defined as follows: if R is an n-ary relational symbol and for each $i \in n$, t_i is a term, then $FreeVar(R(t_0, \ldots, t_{n-1}))$ is $\bigcup_{i\in n} Var(t_i)$; if t is a term and X a second-order variable, then $FreeVar(t{\in}X)$ is the set $Var(t) \cup \{X\}$; if φ_1 and φ_2 are two formulae, then $FreeVar(\varphi_1{\vee}\varphi_2)$ is $FreeVar(\varphi_1) \cup FreeVar(\varphi_2)$; if φ is a formula and x, X are variables (respectively first- and second-order), then $FreeVar(\neg\varphi)$ is $FreeVar(\varphi)$, $FreeVar(\exists x.\ \varphi)$ is $FreeVar(\varphi) \setminus \{x\}$, and $FreeVar(\exists X.\ \varphi)$ is $FreeVar(\varphi) \setminus \{X\}$. A variable that occurs in φ but is not free is called a *bounded* variable of φ. For instance, x is a free variable of the formula $\exists y.\ R(x, y)$, while y is bounded. Notice that, if x is a bounded variable of the formula $\exists x.\ Q(x)$, it is yet a free variable of the formula $P(x) \vee \exists x.\ Q(x)$.

We usually write a formula φ with its free variables in parentheses, meaning $\varphi(x_0, \ldots, x_{p-1}, X_0, \ldots X_{q-1})$ if its free variables are $x_0, \ldots, x_{p-1}, X_0, \ldots X_{q-1}$. We say that a formula is a *sentence* if it is without free variables.

As symbols of a signature Σ have interpretations in a model $\mathcal{M}$, MSO[Σ] formulae have an induced *semantic interpretations* in $\mathcal{M}$, first introduced by Tarski in his *definition of truth* [Tar33][1]. This is now considered as the standard way of defining the *satisfaction* of a formula in a model.

We first have a notion of *valuations*. Let $\mathcal{M}$ be a model of a signature Σ. A *valuation* in $\mathcal{M}$ is a function ρ from a set V of variables that maps first-order variables of V to elements of $\mathcal{M}$ and second-order variables of V to subsets of $\mathcal{M}$. If ρ is a valuation of a set V of variables, x is a first-order variable, and $m \in \mathcal{M}$, then by $\rho[x\mapsto m]$ we denote the valuation of $V \cup \{x\}$ that maps x to m, and that agrees with ρ on all other variables in V. We have a similar definition for $\rho[X\mapsto M]$, with X being a second-order variable and $M \subseteq \mathcal{M}$.

We use these valuations to *evaluate* terms into elements of $\mathcal{M}$: for each term t, each set V of variables such that $Var(t) \subseteq V$, and each valuation of V in $\mathcal{M}$, we define the ρ-evaluation of t, written t^ρ this way: $c^\rho = c^\mathcal{M}$ for every constant symbol c, $x^\rho = \rho(x)$ for every first-order variable, and $f(t_0, \ldots, t_{n-1})^\rho = f^\mathcal{M}\big(t_0^\rho, \ldots, t_{n-1}^\rho\big)$ for every n-ary functional symbol and terms $t_0, \ldots, t_{n-1}$.

Now, let us consider φ a formula over a signature Σ. Let $\mathcal{M}$ be a model of Σ, and let ρ be a valuation of V in $\mathcal{M}$, with V being a set of variables such that $FreeVar(\varphi) \subseteq V$. We inductively define the fact that $\mathcal{M}$ *satisfies* φ up to the valuation ρ, which we denote by $\mathcal{M}, \rho \vDash \varphi$.

- If R is a relational symbol of arity n, and if $t_0, \ldots, t_{n-1}$ are n terms over Σ, then $\mathcal{M}, \rho \vDash R(t_0, \ldots, t_{n-1})$ if the n-tuple $\big\langle t_0^\rho, \ldots, t_{n-1}^\rho \big\rangle$ is in $R^\mathcal{M}$.
- If t is a term, and X a second-order variable, then $\mathcal{M}, \rho \vDash t \in X$ if the element t^ρ is in the set $\rho(X)$.
- If φ is a formula, then $\mathcal{M}, \rho \vDash \neg\varphi$ if M does not satisfy φ up to ρ, a condition which we denote by $\mathcal{M}, \rho \nvDash \varphi$.
- If φ_1 and φ_2 are two formulae, then $\mathcal{M}, \rho \vDash \varphi_1 \vee \varphi_2$ if $\mathcal{M}, \rho \vDash \varphi_1$ or $\mathcal{M}, \rho \vDash \varphi_2$. We remind that this "or" has to be understood in the mathematical inclusive sense, meaning that both conditions can be true.
- If φ is a formula, then $\mathcal{M} \vDash \exists x.\ \varphi$ if there exists some element $m \in \mathcal{M}$ such that $\mathcal{M}, \rho[x\mapsto m] \vDash \varphi$.

[1] To be more precise, Tarski introduced this notion in the context of *First-Order Logic*, which we discuss in Subsection 1.2.2.

– Similarly, if φ is a formula, then $\mathcal{M}, \rho \vDash \exists X.\ \varphi$ if there exists some subset $M \subseteq \mathcal{M}$ such that $\mathcal{M}, \rho[X \mapsto M] \vDash \varphi$.

Starting from now, when we write $\mathcal{M}, \rho \vDash \varphi$, we always assume that ρ is a valuation defined on a set V of variables such that $FreeVar(\varphi) \subseteq V$, even without stating it explicitly.

Sometimes, we will write the valuations inside the parentheses of the formulae. For instance, if $\varphi(x)$ is a formula with a single free first-order variable, and if we actually defined some m as a particular element of $\mathcal{M}$, then we will simply write $\mathcal{M} \vDash \varphi(m)$ for $\mathcal{M}, x \mapsto m \vDash \varphi(x)$. We will make sure along the thesis that no confusion can arise from this.

An first important result is that satisfactions are preserved by isomorphisms of models:

Fact 1.5. *Let $\mathcal{M}$ and $\mathcal{N}$ be two models of a signature Σ. We suppose that there exists an isomorphism ι from $\mathcal{M}$ to $\mathcal{N}$. Then for every* MSO[Σ] *formula φ, and every valuation ρ, we have $\mathcal{M}, \rho \vDash \varphi$ if and only if $\mathcal{N}, \iota \circ \rho \vDash \varphi$.*

This fact is easily shown by induction on φ.

In the thesis, we will simplify our formulae by defining some well-known constructors with the ones already introduced: we define the *conjunction* of φ_1 and φ_2, written $\varphi_1 \wedge \varphi_2$, as $\neg(\neg\varphi_1 \vee \neg\varphi_2)$; we define the *implication* of φ_2 from φ_1, written $\varphi_1 \Longrightarrow \varphi_2$, as $\neg\varphi_1 \vee \varphi_2$; we define the *equivalence* of φ_1 and φ_2, written $\varphi_1 \Longleftrightarrow \varphi_2$, as $(\varphi_1 \Longrightarrow \varphi_2) \wedge (\varphi_2 \Longrightarrow \varphi_1)$; and we write $\forall x.\ \varphi(x)$ for $\neg\exists x.\ \neg\varphi(x)$, $\forall$ being called the *universal quantifier* (we define a similar quantifier for second-order variables). The notions of free variables and satisfaction naturally follow: for instance, $\mathcal{M} \vDash \varphi_1 \wedge \varphi_2$ is true if and only if both $\mathcal{M} \vDash \varphi_1$ and $\mathcal{M} \vDash \varphi_2$ are. Also, we define $\top$ as the formula $\forall x.\ x = x$, it is satisfied by all models. On the contrary, $\bot$, defined as $\neg\top$, is never satisfied. A final shortened notation which we give is the *existential quantifier with unicity*: if $\varphi(x)$ is a formula with a free first-order variable x, then $\exists! x.\ \varphi(x)$ stands for the formula $\exists x.\ \varphi(x) \wedge \forall y.\ \varphi(y) \Longrightarrow x = y$. It expresses the existence of a unique element x that satisfies φ. We naturally have such a similar quantifier for second-order variables.

Moreover, we often simplify the use of the quantifiers: we write $\exists x_0, \ldots, x_{p-1}, X_0, \ldots, X_{q-1}.\ \varphi$ for $\exists x_0. \ldots \exists x_{p-1}. \exists X_0. \ldots \exists X_{q-1}.\ \varphi$, we write $\exists x {\in} X.\ \varphi(x)$ for $\exists x.\ x {\in} X \wedge \varphi(x)$, and similarly for universal quantifiers ($\forall x {\in} X.\ \varphi(x)$ shall be understood as $\forall x.\ x {\in} X \Longrightarrow \varphi(x)$). Similarly as we use the notation $\bigvee_{i \in n} \varphi_i$ (see page 30), we write $\bigwedge_{i \in n} \varphi_i$ for the formula $\varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_{n-1}$. The constructors $\neg$, $\vee$, and $\wedge$ are called the *Boolean operators* of Monadic Second-Order Logic.

Finally, we give a notion of subformulae: if φ is a formula, we inductively define a set $SubForm(\varphi)$ of formulae in the following way: if φ is of the shape $R(t_0, \ldots, t_{n-1})$ or of the shape $t \in X$, then $SubForm(\varphi) = \{\varphi\}$; if φ is of the shape $\neg\psi$, of the shape $\exists x.\ \psi$, or of the shape $\exists X.\ \psi$, then $SubForm(\varphi) = \{\varphi\} \cup SubForm(\psi)$; and if φ is of the shape $\varphi_1 \vee \varphi_2$, then $SubForm(\varphi) = \{\varphi\} \cup SubForm(\varphi_1) \cup SubForm(\varphi_2)$. Similarly for the constructors defined in the previous paragraph. A *subformula* of φ is any formula in $SubForm(\varphi)$.

1.2.2 First-Order Logic and fragments using k variables

In this subsection, we define some fragments of MSO which we also study in the thesis.

Let Σ be a signature. *First-Order Logic* over Σ, written FO[Σ], is the subclass of MSO[Σ] formulae φ that do not contain second-order quantifiers: no subformula of φ is of the shape $\exists X.\ \psi$ nor $\forall X.\ \psi$. We say that φ is an FO[Σ] *formula*, or that it is *first order*. Naturally, First-Order Logic inherits all the notions defined in Subsection 1.2.1 (free variables, satisfaction...), as well as the simplifications of notations. Notice that our definition allows FO[Σ] formulae to have free second-order variables. A complete survey of First-Order Logic can be found in [Rau09].

Many fragments of MSO stronger than FO have been studied. An example of them is *Weak Monadic Second-Order Logic* (denoted by WMSO), see for instance [Rab70]. In this formalism, the second-order quantifiers $\exists^{\text{fin}}$ and $\forall^{\text{fin}}$ only range over finite subsets of the models. Yet, this grammar is not to be worked with in the thesis.

Finally, we introduce another fragment, this time of FO. Let k be some natural number. We define First-Order Logic over Σ with k variables, written FO^k[Σ], as the class of formulae in FO[Σ] that quantify over only k distinct variables, usually $x_0, \ldots, x_{k-1}$. For instance, if P is a predicate symbol of Σ, then the formula $\varphi(x)$ defined as

$$\exists y, z.\ P(x) \wedge P(y) \wedge P(z) \wedge x \neq y \wedge y \neq z \wedge z \neq x$$

is an FO^2[Σ] formula but not an FO^1[Σ] formula, since it quantifies over two variables y and z.

1.2.3 Ehrenfeucht-Fraïssé games

First introduced as the *back-and-forth method* in [Fra55], and then formalised as games in [Ehr61], *Ehrenfeucht-Fraïssé games* are a powerful tool to prove the impossibility to express some properties in given formalisms, such as FO and MSO.

In order to detail the games, we must define two notions: a notion of *submodels*, and a notion of *quantifier depth* of a formula.

Let Σ be a signature, and let us suppose that it contains only relational symbols. Let $\mathcal{M}$ be a model of Σ, and let $\mathcal{M}'$ be a subset of it. It can be seen as a model of Σ, where relational symbols have their interpretations induced from their interpretations in $\mathcal{M}$. Under this assumption, $\mathcal{M}'$ is called a *submodel* of $\mathcal{M}$. If $m_0, \ldots, m_{k-1}$ are elements of $\mathcal{M}$, then a particular submodel is the set $\{m_0, \ldots, m_{k-1}\}$. We denote it by $\mathcal{M}{\restriction}\langle m_0, \ldots, m_{k-1}\rangle$.

A notion of submodels can also be considered when Σ *does* contain functional symbols, but in this thesis, we will work only with signatures that are fully relational, therefore our definition will be sufficient.

Let $\mathcal{M}$ and $\mathcal{N}$ be two models of a relational signature Σ, let $m_0, \ldots, m_{k-1}$ (resp. $n_0, \ldots, n_{k-1}$) be elements of $\mathcal{M}$ (resp. of $\mathcal{N}$). We say that the k-tuples $\langle m_0, \ldots, m_{k-1}\rangle$ and $\langle n_0, \ldots, n_{k-1}\rangle$ are *similar* if the function mapping each m_i to n_i exists (in the sense that $m_i = m_j$ if and only if $n_i = n_j$ for all $i, j \in k$) and is an isomorphism from $\mathcal{M}{\restriction}\langle m_0, \ldots, m_{k-1}\rangle$ to $\mathcal{N}{\restriction}\langle n_0, \ldots, n_{k-1}\rangle$ (we refer to the page 29 for the definition of an isomorphism of models). In particular, the order of the elements is important.

Finally, we define the *quantifier depth* of a formula φ, denoted by $\exists\forall{-}\mathrm{depth}(\varphi)$. Its definition is inductive. It is equal to 0 if φ is atomic, equal to $\exists\forall{-}\mathrm{depth}(\psi)$ if φ is of the shape $\neg\psi$, equal to $\max\big(\exists\forall{-}\mathrm{depth}(\varphi_1), \exists\forall{-}\mathrm{depth}(\varphi_2)\big)$ if φ is of the shape $\varphi_1 \vee \varphi_2$ or of the shape $\varphi_1 \wedge \varphi_2$, and equal to $\exists\forall{-}\mathrm{depth}(\psi) + 1$ if φ is of the shape $\exists x.\ \psi$, $\forall x.\ \psi$, $\exists X.\ \psi$, or $\forall X.\ \psi$. We write $\mathrm{MSO}_d[\Sigma]$ for the class of $\mathrm{MSO}[\Sigma]$ formulae of quantifier depth at most d. More generally, if C is any class of MSO formulae, then we denote by C_d the class of formulae in C of quantifier depth at most d.

Now that we have these notions of quantifier depth, we can define the *Ehrenfeucht-Fraïssé game*, and explain how it is linked to Monadic Second-Order Logic.

The *Ehrenfeucht-Fraïssé game* for $\mathrm{FO}[\Sigma]$ and $(\mathcal{M}, \mathcal{N})$, with d turns, which we denote here by $\mathcal{G}_d^{\mathrm{FO}[\Sigma]}(\mathcal{M}, \mathcal{N})$, is a game played between two players, one called *Spoiler*, the other called *Duplicator*. The rules are the following, on each turn $i \in d$:

- Spoiler selects an element of one of the two models;
- Duplicator also selects an element, but in the other model as Spoiler, more precisely: if Spoiler selected an element m_i (resp. n_i) in $\mathcal{M}$ (resp. in $\mathcal{N}$), then Duplicator has to select an element n_i (resp. m_i) in $\mathcal{N}$ (resp. in $\mathcal{M}$).

After the d turns, Spoiler and Duplicator have defined together two d-tuples $\langle m_0, \ldots, m_{d-1}\rangle$, $\langle n_0, \ldots, n_{d-1}\rangle$, in $\mathcal{M}$ and $\mathcal{N}$ respectively. We say that Duplicator *wins* the game if these d-tuples are similar.

Intuitively, Spoiler wants to prove that the structures of two models are different, while Duplicator tries to show that, even if they are not the same, the structures do have some similarities. We say that the latter player has a *winning strategy* if whatever the selections made by Spoiler, he has a possibility to select elements that ensure him the victory, and we denote by $\mathcal{M} \equiv_d \mathcal{N}$ this assumption (the signature Σ being implicit). It is a standard result that this relation $\equiv_d$ is an equivalence relation over models of Σ, see the definition on page 22[2].

Notice that the rules do not forbid Spoiler to change the model he selects an object from: if he selects an element m_0 in $\mathcal{M}$ on turn 0, he has perfectly the right to select an element n_1 in $\mathcal{N}$ on the next turn.

The following theorem relates this game with the class $\mathrm{FO}_d[\Sigma]$:

Theorem 1.6 ([EF95]). *Let Σ be a relational signature, $\mathcal{M}$, $\mathcal{N}$ two models of Σ, and d some natural number. If $\mathcal{M} \equiv_d \mathcal{N}$, then the two models $\mathcal{M}$ and $\mathcal{N}$ are* indistinguishable *in* $\mathrm{FO}_d[\Sigma]$. *Moreover, if Σ is finite, then the converse is also true.*

In the theorem, $\mathcal{M}$ and $\mathcal{N}$ being *indistinguishable* in a formalism C means that for any sentence φ in C without free variables, we have that $\mathcal{M}$ satisfies φ if and only so does $\mathcal{N}$.

This theorem is very useful to prove that some class $\mathcal{C}$ of models is not definable in $\mathrm{FO}[\Sigma]$: if for all $d \in \mathbb{N}$, we can exhibit two models $\mathcal{M}_d$ and $\mathcal{N}_d$ of Σ such that $\mathcal{M}_d$ is in $\mathcal{C}$, $\mathcal{N}_d$ is not, and $\mathcal{M}_d \equiv_d \mathcal{N}_d$, then we know by Theorem 1.6 that there is no hope of defining a sentence $\varphi_{\mathcal{C}}$ such that a model is in $\mathcal{C}$ if and only if it satisfies $\varphi_{\mathcal{C}}$.

We now present a variation of the game, in which players are allowed to select not only elements, but also subsets of the models. For this, we need a notion of similarity when such subsets are considered.

Each subset M of a model $\mathcal{M}$ can be seen as the interpretation $P^{\mathcal{M}}$ of a new predicate symbol P, with $M = \{x \in \mathcal{M} \mid \mathcal{M} \vDash P(x)\}$. Therefore, if $M_0, \dots M_{l-1}$ are subsets of $\mathcal{M}$, and $N_0, \dots, N_{l-1}$ are subsets of $\mathcal{N}$, $\mathcal{M}$ and $\mathcal{N}$ being two models of Σ, then we can see $\mathcal{M}$ and $\mathcal{N}$ as models of the new signature $\Sigma \sqcup \{P_0, \dots, P_{l-1}\}$, with $M_j = P_j^{\mathcal{M}}$ and $N_j = P_j^{\mathcal{N}}$ for each $j \in \ell$, and say that the tuples $\langle m_0, \dots, m_{k-1}, M_0, \dots, M_{\ell-1}\rangle$ and $\langle n_0, \dots, n_{k-1}, N_0, \dots, N_{\ell-1}\rangle$ are *similar* if the k-tuples $\langle m_0, \dots, m_{k-1}\rangle$ and $\langle n_0, \dots, n_{k-1}\rangle$ are similar when considering the new signature $\Sigma \sqcup \{P_0, \dots, P_{\ell-1}\}$. In other words, $\langle m_0, \dots,$ $m_{k-1}, M_0, \dots, M_{\ell-1}\rangle$ and $\langle n_0, \dots, n_{k-1}, N_0, \dots, N_{\ell-1}\rangle$ are similar if the function $m_i \mapsto n_i$ exists, is an isomorphism from $\mathcal{M}\restriction\langle m_0, \dots, m_{k-1}\rangle$ to $\mathcal{N}\restriction\langle n_0, \dots, n_{k-1}\rangle$, and if moreover we have $m_i \in M_j$ if and only if $n_i \in N_j$ for all $i \in k$, $j \in \ell$.

[2]Technically, we defined equivalence relations on sets, and the models of Σ form a class, but this will not be problematic.

Now, in the new game, denoted by $\mathcal{G}_d^{\mathrm{MSO}[\Sigma]}(\mathcal{M},\mathcal{N})$, as stated above, Spoiler decides on each turn i to select a single element, or a subset, in one of the two models, and Duplicator has to answer with an object of the same type (meaning an element if Spoiler selected an element, and a subset if he selected a subset) and in the other model. After d turns, Spoiler and Duplicator have defined together two d-tuples $\langle m_0,\dots,m_{k-1},M_0,\dots,M_{\ell-1}\rangle$, $\langle n_0,\dots,n_{k-1},N_0,\dots,N_{\ell-1}\rangle$, with $k+\ell=d$. We say that Duplicator *wins* the game if these d-tuples are similar, in the second sense defined above. And, as previously, we say that he has a *winning strategy*, which we write $\mathcal{M}\cong_d\mathcal{N}$, if whatever selections made by Spoiler, he has a possibility to select elements and subsets that ensure him the victory. We write $\mathcal{M}\cong\mathcal{N}$ if $\mathcal{M}\cong_d\mathcal{N}$ for all $d\in\mathbb{N}$. These two relations are also equivalence relations over models of Σ.

In the same way that $\mathcal{G}_d^{\mathrm{FO}[\Sigma]}(\mathcal{M},\mathcal{N})$ is linked with definability in First-Order Logic, this new game $\mathcal{G}_d^{\mathrm{MSO}[\Sigma]}(\mathcal{M},\mathcal{N})$ is linked with definability in Monadic Second-Order Logic:

Theorem 1.7. *Let Σ be a relational signature, $\mathcal{M}$, $\mathcal{N}$ two models of Σ, and d some natural number. If $\mathcal{M}\cong_d\mathcal{N}$, then the models $\mathcal{M}$ and $\mathcal{N}$ are indistinguishable in* $\mathrm{MSO}_d[\Sigma]$*. Moreover, if Σ is finite, then the converse is also true.*

A proof of Theorem 1.7 can be directly adapted from the proof of Theorem 1.6 given in [EF95].

Hence, similarly, this version of the game is very useful to prove that a given class of models is not definable by an MSO[Σ] sentence.

We will use this theorem when working with linear orders. We will also make use of the following proposition, which can be seen as a variant of the *composition method* often used by Shelah [She75]:

Proposition 1.8. *Let λ_1, λ_2, μ_1, and μ_2 be four linear orders, seen as models of the signature $\Sigma=\{<\}$. We suppose that $\lambda_1\cong\mu_1$ and that $\lambda_2\cong\mu_2$. Then $\lambda_1+\lambda_2\cong\mu_1+\mu_2$.*

A proof of this proposition is rather simple, Duplicator's winning strategy for the game $\mathcal{G}_d^{\mathrm{MSO}[\Sigma]}\langle\lambda_1+\lambda_2,\mu_1+\mu_2\rangle$ being basically the "concatenation" of his winning strategies for $\mathcal{G}_d^{\mathrm{MSO}[\Sigma]}\langle\lambda_1,\mu_1\rangle$ and $\mathcal{G}_d^{\mathrm{MSO}[\Sigma]}\langle\lambda_2,\mu_2\rangle$.

Finally, we give a last variation of the original Ehrenfeucht-Fraïssé game: the *first-order variation with k tokens*, written $\mathcal{G}_d^{\mathrm{FO}^k[\Sigma]}\langle\mathcal{M},\mathcal{N}\rangle$.

In this version, we consider, as previously, two models $\mathcal{M}$ and $\mathcal{N}$ of a relational signature Σ, but, this time, $2k$ tokens are given between the two players: two tokens with the number 0 on them, two tokens with the number 1 on them, and so on up to the number $k-1$.

The tokens with the number j on them are called the two *j-tokens*. At the beginning of the game, all the tokens are unplaced. Then, at each turn, Spoiler takes one token (let us say that it is a j-token) and places it on one element of one of the models. Notice that, before that, this token could be unplaced, as it could already be placed somewhere on one of the models. After this, Duplicator takes the other j-token, and places it on an element of the other model.

With these rules, we ensure that for each $j \in k$, either the two j-tokens are unplaced, or they are both placed on the two different models. At each turn $i \in d$, we denote by E_i the set of numbers j in $\{0, \ldots, k-1\}$ such that, after turn i, the two j-tokens have been placed on the models. Hence, after each turn i, the two players have defined two $|E_i|$-tuples: one tuple $(m_j)_{j \in E_i}$ of elements of $\mathcal{M}$, and one tuple $(n_j)_{j \in E_i}$ of elements of $\mathcal{N}$. We say that Duplicator wins this game if after each turn $i \in d$, these two tuples are similar (in the first sense, since no subsets are involved here). We write $\mathcal{M} \equiv_d^k \mathcal{N}$ if Duplicator has a *winning strategy* (the notion of a *winning strategy* being similar as the ones above). Finally, we write $\mathcal{M} \equiv^k \mathcal{N}$ if $\mathcal{M} \equiv_d^k \mathcal{N}$ for every $d \in \mathbb{N}$. Like the previous relations, $\equiv_d^k$ and $\equiv^k$ are equivalence relations over models of Σ.

Now, we can state the following theorem linking this variation and the class $\mathrm{FO}^k[\Sigma]$:

Theorem 1.9 ([Imm99])**.** *Let Σ be a relational signature, $\mathcal{M}$, $\mathcal{N}$ two models of Σ, and k, d, two natural numbers. If $\mathcal{M} \equiv_d^k \mathcal{N}$, then the models $\mathcal{M}$ and $\mathcal{N}$ are indistinguishable in $\mathrm{FO}_d^k[\Sigma]$. Moreover, if Σ is finite, then the converse is also true.*

There are many more variations of the original Ehrenfeucht-Fraïssé game. In fact, there are as many variations as there are formalisms. However, only the ones introduced in this section will be used later in the thesis.

1.2.4 Regular and First-Order Logic definable languages of words

In Subsections 1.2.1 and 1.2.2, we introduced MSO and some of its fragments, in the general sense. Here, we focus our study of these formalisms in the case of words.

Let $\mathbb{A}$ be an alphabet. We consider the signature $\mathbb{A} \sqcup \{<\}$, where: for each $a \in \mathbb{A}$, a is a predicate (sometimes also denoted by P_a in the literature), and $<$ is a binary relational symbol. A *model* of $\mathbb{A} \sqcup \{<\}$ is any word w over $\mathbb{A}$: the elements of the model are the positions of $\mathrm{Dom}(w)$, the interpretation $<^w$ of $<$ in w is naturally the order of $\mathrm{Dom}(w)$, and the interpretation a^w of the predicate symbol a, for $a \in \mathbb{A}$, is the set $\{x \in \mathrm{Dom}(w) \mid w(x) = a\}$ of positions labelled by a. Since we always suppose the equality symbol is in the signature (see our remark about this convention on page 29), we can also use the non-strict order relation $\leq$ when needed. Finally, we allow some natural shortcuts in the grammar: for instance $\exists y{<}x.\ \varphi(x,y)$ is a shortened notation for the formula $\exists y.\ y{<}x \wedge \varphi(x,y)$.

If φ is an MSO$[\mathbb{A},<]$ formula, with $\mathbb{A}$ being an alphabet, and if λ is a linear order, then we denote by $\mathcal{L}^\lambda(\varphi)$ the λ-language $\{w \in \mathbb{A}^\lambda \mid w \vDash \varphi\}$, and we call it the *language* of φ over λ. We also define $\mathcal{L}^+(\varphi) := \{w \in \mathbb{A}^+ \mid w \vDash \varphi\}$ and $\mathcal{L}^\circ(\varphi) := \{w \in \mathbb{A}^\circ \mid w \vDash \varphi\}$, the *languages* of φ, respectively in the cases of finite and of countable words. In the thesis, we will often simply write $\mathcal{L}(\varphi)$ for all these languages, the context making it clear whether we speak of finite words, of countable words, or of words over a particular linear order λ. A language over $\mathbb{A}$ is called *regular* if it is $\mathcal{L}(\varphi)$ for some MSO$[\mathbb{A},<]$ formula φ. In this case we say that it is *defined* by φ. We denote by **MSO**$[\mathbb{A},<]$, with bold letters, the class of regular languages over the alphabet $\mathbb{A}$, and, more generally, **MSO**$[<]$ is the class of regular languages over any alphabet.

We extend theses notions of definability to subclasses of MSO$[\mathbb{A},<]$ formulae. In general, if C is a class of formulae (for instance the class FO$[\mathbb{A},<]$), then we say that a language L of words is *definable* in C if it is $\mathcal{L}(\varphi)$ for some formula φ in C, and we denote by **C**, with bold letters, the class of these languages of words definable in C. Once again, the context making it clear whether we consider these languages over finite words, over countable words, or over a fixed linear order.

We also extend these notions to relations: if R is a binary relation over a linear order λ, we say that it is *regular* (resp. definable in C) if there exists some MSO$[\mathbb{A},<]$ formula (resp. some formula in C) φ such that $R = \big\{\langle x,y\rangle \in \lambda \times \lambda \mid \lambda \vDash \varphi(x,y)\big\}$. We do the same for predicates, for n-ary relations, for relations between elements and subsets, etc.

In the case of finite words, we will often consider another relation between positions: the *successor relation*, denoted by the bold letter s. If the domain is the natural number n, and if $x,y \in n$, then $\mathsf{s}(x,y)$ expresses that $y = x{+}1$. In the literature, s is also often defined as a function from n to itself, with each element $x \neq n{-}1$ being mapped to $x{+}1$ and $n{-}1$ being mapped to itself. When at least two variables are allowed in the formulae, these two definitions are equivalent, and, in the thesis, we choose the former option, in order to have a fully relational signature $\mathbb{A} \sqcup \{\mathsf{s}\}$. Naturally, this successor relation can also be introduced when considering ω-words.

We will now focus on the context of finite words, and state some known results of inclusions and equalities between the different fragments (**MSO**, **FO**, $\mathbf{FO}^k \ldots$). The majority of the inclusions, but not all of them, are naturally deduced from the richness of the formalisms: if a first formalism is richer than a second, then any language definable in the latter is naturally also definable in the former.

A first thing to notice is that using the successor relation s or the order relation does not change the expressive power of MSO:

Proposition 1.10. *In the case of finite words, the following three classes have exactly the same languages:* $\mathbf{MSO}[<]$*,* $\mathbf{MSO}[\mathsf{s}]$*, and* $\mathbf{MSO}[<,\mathsf{s}]$*.*

Proof. First, if x and y are two positions of a word, then the property $\mathsf{s}(x,y)$ can be defined by the formula $x{<}y \wedge \forall z.\ x{<}z \Longrightarrow y{\leq}z$, which is an MSO[<] formula. Reciprocally, $x \leq y$ can be defined by the MSO[s] formula saying that any subset of the domain containing x and stable by s necessarily contains y: $\forall X.\ \big[x \in X \wedge \forall x_0, x_1.\ \big(x_0 \in X \wedge \mathsf{s}(x_0,x_1)\big) \Longrightarrow x_1 \in X\big] \Longrightarrow y \in X$. $\square$

Hence, in Monadic Second-Order Logic, in the case of finite words, the question of using one relation or the other is simply a matter of convenience: we will simply write $\mathbf{MSO}[<]$ for the class, but we will also use s whenever we need it. Notice that our definition of the successor relation, using $<$, does not involve any second-order variable, hence, it can also be used to prove the inclusion $\mathbf{FO}[\mathsf{s}] \subseteq \mathbf{FO}[<]$:

Corollary 1.11. *In the case of finite words, we have* $\mathbf{FO}[\mathsf{s}] \subseteq \mathbf{FO}[<] = \mathbf{FO}[<,\mathsf{s}]$*.*

Hence, we can now simply write $\mathbf{FO}[<]$ for $\mathbf{FO}[<,\mathsf{s}]$, even if we use the relation s in our formulae when it is useful.

Now, we show some results for the classes $\mathbf{FO}^k[\Sigma]$. First, it is evident that, independently of the signature Σ, $\mathbf{FO}^k[\Sigma] \subseteq \mathbf{FO}^{k'}[\Sigma]$ as soon as $k \leq k'$. An important result was shown in Kamp's PhD thesis [Kam68]. It states that we reach the whole class $\mathbf{FO}[<]$ with only 3 variables:

Theorem 1.12 ([Kam68])**.** *In the case of finite words, we have* $\mathbf{FO}^3[<] = \mathbf{FO}^3[<,\mathsf{s}] = \mathbf{FO}[<]$*.*

All these inclusions and equalities leads to the following graph of class of languages of finite words, each arrow representing an inclusion:

On the last line of the graph, the classes $\mathbf{FO}^k[\,]$ and $\mathbf{FO}[\,]$ are the classes of languages defined by formulae using neither the order $<$, nor the successor symbol s, but only the predicate symbols inherited from the alphabets. We will tell more about these languages in Chapter 2.

In fact, it happens that all the inclusions of Figure 1.1 are strict, as we are going to see later in the thesis.

$$
\begin{tikzcd}
 & & & & & \mathbf{MSO}[<] \\
\mathbf{FO}^2[<] \arrow[r] & \mathbf{FO}^2[<,\mathsf{s}] \arrow[rrrr] & & & & \mathbf{FO}[<] \arrow[u] \\
\mathbf{FO}^2[<] \cap \mathbf{FO}^2[\mathsf{s}] \arrow[u] \arrow[r] & \mathbf{FO}^2[\mathsf{s}] \arrow[u] \arrow[r] & \cdots \arrow[r] & \mathbf{FO}^k[\mathsf{s}] \arrow[r] & \cdots \arrow[r] & \mathbf{FO}[\mathsf{s}] \arrow[u] \\
\mathbf{FO}^2[\,] \arrow[u] \arrow[r] & \cdots \arrow[rr] & & \mathbf{FO}^k[\,] \arrow[u] \arrow[r] & \cdots \arrow[r] & \mathbf{FO}[\,] \arrow[u]
\end{tikzcd}
$$

Figure 1.1: The main classes of regular languages of finite words studied in the thesis.

The semantic interpretation of the constructors $\vee$, $\wedge$, and $\neg$ (see pages 31 and 32) make it clear that all of these classes are *closed under Boolean operations*: if the languages L_1 and L_2 are in one of these classes, say $\mathbf{C}$, then the languages $L_1 \cup L_2$, $L_1 \cap L_2$, and L_1^{c} are also in $\mathbf{C}$.

Additionally, these classes have the other property of being closed under *extensions of alphabets*: if L is a language of finite words over some alphabet $\mathbb{A}_1$ (*i.e.* we consider the language $\langle L, \mathbb{A}_1 \rangle$, see page 24) that is in one of these classes, say $\mathbf{C}$, and if $\mathbb{A}_2$ is an alphabet such that $\mathbb{A}_1 \subseteq \mathbb{A}_2$, then the language $\langle L, \mathbb{A}_2 \rangle$, is also in $\mathbf{C}$. Indeed, if φ is a formula in C that defines $\langle L, \mathbb{A}_1 \rangle$, then the formula $\varphi \wedge \forall x. \bigvee_{a \in \mathbb{A}_1} a(x)$ is also in C and clearly defines $\langle L, \mathbb{A}_2 \rangle$.

Notice that these properties of closure under Boolean operations and extensions of alphabets hold not only for the classes of Fig 1.1 over finite words, but also for classes such as $\mathbf{MSO}[<]$ over countable words, or the class of languages of countable words definable by a formula in WMSO (see the definition on page 33).

A class of languages that has some natural closure properties such as these, is called *robust*. There is no formal definition of robustness, since many different closures can be considered, as we shall see in the very next section of this chapter. Most well-studied classes happen to have these closures under Boolean operations and extensions of alphabets.

1.3 Algebra

Regular languages are often studied through *algebraic objects*, *i.e.* sets with operations satisfying some properties. This usage is convenient since, as we will show it in this section, a lot of theorems highlight that regular languages, even if infinite, admit finite representations.

In Subsection 1.3.1, we show the link between *semigroups* and languages over finite words. In particular, we present the Myhill-Nerode theorem, which states that regular languages of finite words can be represented with finite semigroups. In Subsection 1.3.2, we see that natural closures of classes of languages of finite words, such as closure under Boolean operations, correspond to particular closures of classes of finite semigroups, called *varieties of finite semigroups*. Finally, in Subsection 1.3.3, we introduce $\circ$-*semigroups* and $\circ$-*algebrae* (the symbol $\circ$ being pronounced "circle"), which can be seen as a generalisation of semigroups, and which are useful in the study of languages of countable words.

1.3.1 Semigroups and their connections to classes of languages of finite words

A *law* over a set S is any function from S^2 to S, we generally denote it by the symbol $*$, in the thesis. It is called *associative* if it satisfies $(r * s) * t = r * (s * t)$ for all $r, s, t \in S$. A non-empty set S with an associative law $*$ is called a *semigroup*. The law is also called the *product* of S. Because it is associative, if $s_0, \ldots, s_{n-1}$ are n elements of S, we can simply write $s_0 * s_1 * s_2 * \cdots * s_{n-2} * s_{n-1}$ for $(\cdots((s_0 * s_1) * s_2) * \cdots * s_{n-2}) * s_{n-1}$, *i.e.* without taking care of the parentheses. If $s \in S$ and n is a positive natural number, then we write s^n for the product $s * s * \cdots * s$, with s appearing exactly n times.

If S_1 and S_2 are two semigroups, then we denote by $S_1 \times S_2$ the *product semigroup* of S_1 and S_2, with its law defined by $\langle s_1, s_2\rangle * \langle t_1, t_2\rangle = \langle s_1 * t_1, s_2 * t_2\rangle$ for $s_1, t_1 \in S_1$ and $s_2, t_2 \in S_2$. A *subsemigroup* of a semigroup S is a subset T of S that is *stable* by the law of S, *i.e.* that is such that for all $s, t \in S$, if $s, t \in T$ then also $s * t \in T$. It is immediate to see that in this case, T with $*$ restricted to it is a semigroup. Finally, a *homomorphism* h between two semigroups S_1 and S_2 is a function from S_1 to S_2 that preserves the associative law, meaning that for all $s, t \in S_1$, we have $h(s * t) = h(s) * h(t)$. Notice that the symbol $*$ on the left-hand side of the equality refers to the law of S_1, while the one on the right-hand side refers to the law of S_2. In this case, the subset $h(S_1) \subseteq S_2$ of images of S_1 under h is a subsemigroup of S_2. The composition of two homomorphisms is also a homomorphism. Moreover, a homomorphism is called an *isomorphism* (of semigroups) if it is also bijective. The inverse function of an isomorphism is an isomorphism, and the composition of two isomorphisms is also an isomorphism.

For any alphabet $\mathbb{A}$, the set $\mathbb{A}^+$ of finite words over $\mathbb{A}$, with the concatenation operation, is a semigroup, called the *free semigroup* generated by $\mathbb{A}$. A language $L \subseteq \mathbb{A}^+$ is *recognised* by a semigroup S if there exists a homomorphism h between $\mathbb{A}^+$ and S such that $L =$

$h^{-1}(H) := \{w \in \mathbb{A}^+ \mid h(w) \in H\}$ for some $H \subseteq S$, or, equivalently, if $L = h^{-1}(h(L))$. We also say that the language L is recognised by S *via* h. Notice that, if h is any homomorphism from $\mathbb{A}^+$ to a semigroup S, then for every natural number n and all letters $a_0, \ldots, a_{n-1}$, we have $h(a_0 \cdot a_1 \cdots a_{n-1}) = h(a_0) * h(a_1) * \cdots * h(a_{n-1})$. This means that the image of any word of $\mathbb{A}^+$ by h is induced from the images of the letters, and, therefore, throughout the thesis, we often define such a homomorphism simply by its values on the letters. In particular, a homomorphism between $\mathbb{A}^+$ to $\mathbb{B}^+$ is induced from a function from letters of $\mathbb{A}$ to finite words over $\mathbb{B}$.

Let $L \subseteq \mathbb{A}^+$ be a language of finite words, over an alphabet $\mathbb{A}$. We define the Myrill-Nerode relation $\equiv_L$ over $\mathbb{A}^+$, first introduced in [Ner58], by $u \equiv_L v$ if for all $w_1, w_2 \in \mathbb{A}^*$, we have $w_1 \cdot u \cdot w_2 \in L$ if and only if $w_1 \cdot v \cdot w_2 \in L$. We verify easily that $\equiv_L$ is an equivalence relation on $\mathbb{A}^+$ (see the definition on page 22).

Moreover, $\equiv_L$ is a *congruence*: it has the property that for all u_1, u_2, v_1, and v_2 in $\mathbb{A}^+$, if $u_1 \equiv_L v_1$ and $u_2 \equiv_L v_2$ then $u_1 \cdot u_2 \equiv_L v_1 \cdot v_2$. This implies that the set $\mathbb{A}^+/\equiv_L$ of the equivalence classes of $\equiv_L$ can be given the product defined by $[u]_{\equiv_L} * [v]_{\equiv_L} = [u \cdot v]_{\equiv_L}$, to form a semigroup. This semigroup recognises L via the natural homomorphism $u \mapsto [u]_{\equiv_L}$, and we call it the *syntactic semigroup* of L. In this thesis, we denote it by S_L, and we denote the homomorphism from $\mathbb{A}^+$ to S_L defined above by h_{S_L}.

Proposition 1.13. *Let L be a language of finite words, over an alphabet $\mathbb{A}$. Then S_L is the smallest semigroup recognising L, in the sense that for every semigroup S recognising L via some homomorphism $h\colon \mathbb{A}^+ \to S$, there exists a surjective homomorphism f from S to S_L such that $h_{S_L} = f \circ h$.*

All these notions and theorems can be found in Chapter 3 of [HMU06], but in terms of *deterministic finite automata*, similar tools useful to the study regular languages.

The Myhill-Nerode Theorem makes the link between regular languages and finite syntactic semigroups:

Theorem 1.14 ([Ner58])**.** *Let L be a language of finite words over an alphabet $\mathbb{A}$. Then L is regular if and only if its syntactic semigroup is finite.*

This theorem is of high importance, as it tells us that working with regular languages, which are generally infinite, is equivalent to working with finite semigroups. Moreover, the translations from formulae to finite semigroups, and inversely, are computable:

Proposition 1.15. *There exists an algorithm that inputs an* $\mathrm{MSO}[\mathbb{A},<]$ *sentence* φ*,* $\mathbb{A}$ *being any alphabet, and outputs the (finite) syntactic semigroup* S_L *of the language* $L = \mathcal{L}(\varphi)$*, the syntactic homomorphism* h_L *(given as a function from* $\mathbb{A}$ *to* S_L*), and the subset* $h_{S_L}(L) \subseteq S_L$*.*

Reciprocally, there exists an algorithm that inputs a finite semigroup S*, a homomorphism from* $\mathbb{A}^+$ *to* S *(given as a function from* $\mathbb{A}$ *to* S*), and a subset* $H \subseteq S$*, and outputs an* $\mathrm{MSO}[\mathbb{A},<]$ *sentence* φ *satisfying* $\mathcal{L}(\varphi) = h^{-1}(H)$*.*

These two constructions can also be found in [HMU06, Chapter 3]. The corollary of this result is that it is decidable to test if a given formula is a tautology:

Corollary 1.16. *There exists an algorithm that inputs an* $\mathrm{MSO}[\mathbb{A},<]$ *sentence* φ*, with* $\mathbb{A}$ *being any alphabet, and outputs* **YES** *if the language* $\mathcal{L}(\varphi)$ *is the full language* $\mathbb{A}^+$*, and* **NO** *if it is not.*

Proof. It suffices to notice that the syntactic semigroup of $\mathbb{A}^+$ is composed of a unique element. Let $L = \mathcal{L}(\varphi)$. Considering this, in order to test if L is the full language, it suffices to compute the syntactic semigroup S_L of L and $H = h_{S_L}(L) \subseteq S_L$ (via the algorithm of Proposition 1.15), and to check if S_L is a singleton and if $H = S_L$. $\square$

From this, we know that it is also decidable whether a regular language is included in another, a result which we will use later:

Corollary 1.17. *There exists an algorithm that inputs two* $\mathrm{MSO}[\mathbb{A},<]$ *sentences* φ *and* ψ*, with* $\mathbb{A}$ *being any alphabet, and outputs* **YES** *if the language* $\mathcal{L}(\varphi)$ *is included in the language* $\mathcal{L}(\psi)$*, and* **NO** *if it is not.*

Proof. It suffices to test if the formula $\varphi \implies \psi$ defines the full language $\mathbb{A}^+$, via the algorithm of Corollary 1.16. $\square$

Now, an important notion in the study of semigroups is the notion of *idempotence*: let S be a semigroup, we say that $e \in S$ is *idempotent* if $e^2 = e$, and therefore if $e^n = e$ for all $n \geq 1$. A known result states that, in finite semigroups, exponentiation eventually produces idempotent elements:

Proposition 1.18. *Let* S *be a finite semigroup. Then there exists some positive natural number* n *such that for all* $s \in S$*, the element* s^n *is idempotent. Moreover, there exists an algorithm that inputs* S *and outputs such an* n*.*

A proof of a weaker proposition, where the natural number $n(s)$ also depends on the element s, can be found for example in [Whi78, Chapter 5, Exercise 4]. Then, n obtained as the product of all these individual $n(s)$ satisfies the wanted property. Notice that for all $s \in S$, s^n is the unique idempotent power of s, for if $p, q \geq 1$ are such that s^p and s^q are both idempotent, we have $s^p = (s^p)^q = s^{p \times q} = (s^q)^p = s^q$. Therefore, it makes sense to denote by $\sharp(S)$ the least natural positive number n such that s^n is idempotent for all $s \in S$.

In the literature, we sometimes meet the notation $\omega(S)$ for $\sharp(S)$, but we rather avoid it here, since we already defined the operation $(\cdot)^\omega$ in Subsection 1.1.3. We will also simply write $\sharp$ for $\sharp(S)$.

A finite semigroup S such that for every $s \in S$, $s^\sharp * s = s^\sharp$ is called *aperiodic*. Schützenberger's theorem gives a characterisation of languages definable in FO[<] using aperiodic semigroups:

Theorem 1.19 ([Sch65])**.** *Let $L \subseteq \mathbb{A}^+$ be a language of finite words, over an alphabet $\mathbb{A}$. Then L is definable in* $\mathrm{FO}[\mathbb{A}, <]$ *if and only if its syntactic semigroup S_L is finite and aperiodic.*

Another characterisation which we will use in the thesis is the one concerning **FO**[s]:

Theorem 1.20 ([BP91, Theorem 4.3, together with Corollary 3.8])**.** *Let $L \subseteq \mathbb{A}^+$ be a language of finite words, over an alphabet $\mathbb{A}$. Then L is definable in* $\mathrm{FO}[\mathbb{A}, \mathsf{s}]$ *if and only if its syntactic semigroup S_L is finite, aperiodic, and satisfies the following equation: $e{*}r{*}f{*}s{*}e{*}t{*}f = e{*}t{*}f{*}s{*}e{*}r{*}f$ for all $e, f, r, s, t \in S$, with e and f idempotent.*

Later in the thesis, we will use the following implication of this theorem:

Corollary 1.21. *Let $L \subseteq \mathbb{A}^+$ be a language of finite words over an alphabet $\mathbb{A}$. If it is definable in* $\mathrm{FO}[\mathbb{A}, \mathsf{s}]$*, then, for all $s, t, e \in S_L$, the syntactic semigroup of L, with e idempotent, we have $e{*}s{*}e{*}t{*}e = e{*}t{*}e{*}s{*}e$.*

Proof. It suffices to consider the case $e = f = r$ in the equation of Theorem 1.20. $\square$

In the thesis, we will also use the algebraic characterisation of the class $\mathbf{FO}^2[<, \mathsf{s}]$:

Theorem 1.22 ([PS16, Theorem 3])**.** *Let $L \subseteq \mathbb{A}^+$ be a language of finite words, over an alphabet $\mathbb{A}$. Then L is definable in* $\mathrm{FO}^2[\mathbb{A}, <, \mathsf{s}]$ *if and only if its syntactic semigroup S_L is finite, aperiodic, and satisfies the following equation: $(e * s * e * t * e)^\sharp * t * (e * s * e * t * e)^\sharp = (e * s * e * t * e)^\sharp$ for all $s, t, e \in S$, with e idempotent.*

Similar algebraic characterisations are also known for other fragments of First-Order Logic. For instance, see [TW98, Theorem 6], or [WI07, Fact 1.1], for an algebraic characterisation of $\mathbf{FO}^2[<]$. However, we will not use them in the thesis.

From these algebraic characterisations, we can deduce that some inclusions of classes depicted on Figure 1.1 of the previous section are strict. For instance, $\mathbf{FO}[\mathsf{s}] \subset \mathbf{FO}[<] \subset \mathbf{MSO}[<]$.

1.3.2 Varieties of finite semigroups and varieties of languages

At the end of Section 1.2, we gave an informal notion of robustness: a class of languages is said robust if it has some convenient closure properties, such as closure under Boolean operations. We explicit this idea here and state an important result, the theorem of Eilenberg, highlighting a relation between closure properties for regular languages and closure properties for finite semigroups.

First, let us notice that, thanks to the theorem of Myhill-Nerode, we have a natural correspondence between classes of regular languages and classes of finite semigroups. If $\mathbf{C}$ is a class of languages of finite words, then we can define $\mathbf{Synt}(\mathbf{C})$ as the class of syntactic semigroups S_L of the languages L in $\mathbf{C}$. In this case, if $\mathbf{C}$ contains only regular languages, meaning that all the languages in $\mathbf{C}$ are regular, then by Theorem 1.14, we know that each semigroup in $\mathbf{Synt}(\mathbf{C})$ is finite. Reciprocally, if $\mathbf{V}$ is a class of semigroups, then $\mathbf{Reco}(\mathbf{V})$ is the class of languages recognised by some semigroup in $\mathbf{V}$. Here again, if $\mathbf{V}$ is a class of finite semigroups, we know by the Myhill-Nerode theorem that $\mathbf{Reco}(\mathbf{V})$ contains only regular languages.

Now, we specify the closure properties we want our classes of languages to satisfy. We say that a class $\mathbf{C}$ of regular languages is a *variety of languages* if it satisfies the three following closure properties:

- closure under Boolean operations: if $L_1, L_2 \in \mathbf{C}$, over some alphabet $\mathbb{A}$, then also the languages L_1^c and $L_1 \cup L_2$ are in $\mathbf{C}$;
- *closure under quotients*: if $L \in \mathbf{C}$, over some alphabet $\mathbb{A}$, and $u \in \mathbb{A}^+$, then the two *quotient languages* $u^{-1} \cdot L := \{v \in \mathbb{A}^+ \mid u \cdot v \in L\}$ and $L \cdot u^{-1} := \{v \in \mathbb{A}^+ \mid v \cdot u \in L\}$ also are in $\mathbf{C}$;
- *closure under preimages under homomorphisms*: if $L \in \mathbf{C}$, over an alphabet $\mathbb{B}$, and if h is a homomorphism from $\mathbb{A}^+$ to $\mathbb{B}^+$, then the language $h^{-1}(L)$, over the alphabet $\mathbb{A}$, also is in $\mathbf{C}$.

Concerning the second closure property of this definition, notice indeed that the two quotient languages are regular when L is: if $L \subseteq \mathbb{A}^+$ is recognised by some finite semigroup S via a homomorphism h, with $H \subseteq S$ such that $L = h^{-1}(H)$, and if $u \in \mathbb{A}^+$, then we define $H' = h(u)^{-1} * H := \{x \in S \mid h(u) * x \in H\}$, and we obtain, for $v \in \mathbb{A}^+$, the equivalences $v \in h^{-1}(H')$ iff $h(u) * h(v) \in H$ iff $h(u \cdot v) \in H$ iff $u \cdot v \in L$ (by the assumption on H) iff $v \in u^{-1} \cdot L$, and therefore $u^{-1} \cdot L = h^{-1}(H')$ is also recognised by S via h, and similarly for $L \cdot u^{-1}$.

The same way, if $L \subseteq \mathbb{B}^+$ is recognised by some finite semigroup S, via some homomorphism g, and if $h \colon \mathbb{A}^+ \to \mathbb{B}^+$ is a homomorphism, then $h^{-1}(L)$ is also recognised by S, via the homomorphism $g \circ h$, hence it is regular.

With these two remarks, we can conclude that the class of regular languages is an example of varieties of languages:

Proposition 1.23. **MSO**$[<]$*, the class of regular languages of finite words, is a variety of languages.*

In a second step, we define natural closure properties for a class $\mathbf{V}$ of finite semigroups. We say that it is a *variety of finite semigroups* (or more simply, a *v.f.s.*) if it satisfies these three closure properties:

- *closure under products*: if S_1 and S_2 are in $\mathbf{V}$, then also the product semigroup $S_1 \times S_2$ is in $\mathbf{V}$;
- *closure under subsemigroups*: if S is in $\mathbf{V}$ and T is a subsemigroup of S, then also T is in $\mathbf{V}$;
- *closure under images under homomorphisms*: if S_1 is in $\mathbf{V}$ and h is a homomorphism from S_1 to some semigroup S_2, then also the image semigroup $h(S_1)$ is in $\mathbf{V}$.

In the literature, a variety of finite semigroups is sometimes also called a *pseudo-variety of semigroups*. Most classes, if not all classes of finite semigroups considered in this thesis are varieties of finite semigroups, which justifies why we denote them by the bold letter $\mathbf{V}$.

It happens that these closure properties of v.f.s. correspond to the closure properties of varieties of languages, as states the following theorem:

Theorem 1.24 ([Eil74])**.** *The function* **Synt** *makes a bijection from the varieties of languages to the varieties of finite semigroups, and the function* **Reco** *is its inverse.*[3]

Knowing this one-to-one correspondence, we will often identify a variety $\mathbf{C}$ of languages with its corresponding v.f.s. $\mathbf{Synt}(\mathbf{C})$. We can verify that **MSO**$[<]$, **FO**$[<]$, **FO**$[\mathsf{s}]$, and

[3]Here again, we technically defined functions and bijections only for sets, and not classes, but this will not be problematic.

more generally all the classes of languages introduced in Section 1.2 are varieties of languages:

Proposition 1.25. *The classes of languages of finite words pictured on Figure 1.1, on page 40, are all varieties of languages.*

This proposition is proven in Appendix A.

1.3.3 $\circ$-semigroups and $\circ$-algebrae

We saw in Subsection 1.3.1 that languages of finite words can be studied through semigroups. Following the work of Carton, Colcombet, and Puppis in [CCP18], we show here how this notion can be generalised to $\circ$-*semigroups* and $\circ$-*algebrae*, which will be useful in the study of languages of countable words.

Let S be a non-empty set. We decide to treat it as an alphabet and to consider countable words over it (see Paragraph 1.1.3). This is one of the very few cases in the thesis where we allow an alphabet to be infinite. A *generalised product* over S is an operation $\pi\colon S^\circ \to S$ such that $\pi(s) = s$ for every element $s \in S$, and that moreover satisfies the following property of *generalised associativity*: if I is a countable linearly ordered set, and if for each $i \in I$, γ_i is a countable word over S, then we must have

$$\pi\left(\sum_{i\in I} \pi(\gamma_i)\right) = \pi\left(\sum_{i\in I} \gamma_i\right), \tag{1.1}$$

where the left-hand side sum ranges over single-letter words $\pi(\gamma_i)$; and the right-hand side sum is just the concatenation of all the words γ_i.

We call a $\circ$-*semigroup* (pronounced "circle-semigroup") any set with a generalised product. A remark could be made on the fact that we denote $\circ$-semigroups by the same letter we use to denote semigroups, *i.e.* S, but the context will never leave doubt on which of the two notions we are working with.

Let us define homomorphisms for $\circ$-semigroups: a function h from a $\circ$-semigroup S_1 to a $\circ$-semigroup S_2 is a *homomorphism* if for every word $\gamma \in S_1^\circ$, we have $h(\pi(\gamma)) = \pi(h(\gamma))$, where $h(\gamma)$ has to be understood as the generalised concatenation $\Sigma_{x\in\mathrm{Dom}(\gamma)} h(\gamma(x))$. Here also, notice that the product π on the left-hand side of the equality refers to the generalised product of S_1, while the one on the right-hand side refers to the generalised product of S_2.

For any alphabet $\mathbb{A}$, the set $\mathbb{A}^\circ$ of countable words over $\mathbb{A}$, with the generalised concatenation operation $(\gamma_i)_{i\in I} \mapsto \Sigma_{i\in I}\gamma_i$, is a $\circ$-semigroup, called the *free* $\circ$-*semigroup* generated by $\mathbb{A}$.

As over finite words, we have a notion of recognition by $\circ$-semigroups: a language $L \subseteq \mathbb{A}^\circ$ is *recognised* by a $\circ$-semigroup S via a homomorphism $h\colon \mathbb{A}^\circ \to S$ if $L = h^{-1}(H)$ for some $H \subseteq S$ (or, equivalently, if $L = h^{-1}(h(L))$). As for finite words, regularity can be expressed in terms of recognition by finite algebraic objects:

Theorem 1.26 ([CCP18, Theorems 27 and 30])**.** *Let $L \subseteq \mathbb{A}^\circ$ be a language of countable words over $\mathbb{A}$. Then L is regular if and only if it is recognised by some finite $\circ$-semigroup.*

The inherent difficulty with $\circ$-semigroups is that they are not finite objects: even if S is finite, there is no way to represent every countable word in S° and therefore no way either to represent the generalised product π, in general. This justifies the introduction of similar algebraic objects, but admitting finite representations: $\circ$-*algebrae*.

A $\circ$-*algebra* (which we pronounce "circle-algebra") is a set S with four operations: a law $*\colon S^2 \to S$, two operations $\tau, \tau^\star\colon S \to S$, and an operation $\kappa\colon \mathcal{P}(S)\backslash\{\emptyset\} \to S$. These operations must fulfil a certain number of axioms:

- S with the law $*$ is a semigroup, meaning that $*$ is associative;
- for all $s, t \in S$, $(s * t)^\tau = s * (t * s)^\tau$ and $(s^n)^\tau = s^\tau$ for any natural number $n \geq 1$;
- symmetrically for $\tau^\star$: for all $s, t \in S$, $(s * t)^{\tau^\star} = (t * s)^{\tau^\star} * t$ and $(s^n)^{\tau^\star} = s^{\tau^\star}$ for any $n \geq 1$;
- for each non-empty subset K of S, for all $r \in K, K' \subseteq K$, and each non-empty subset K'' of $\bigcup_{s,t\in K}\{K^\kappa, s{*}K^\kappa, K^\kappa{*}t, s{*}K^\kappa{*}t\}$, we have $K^\kappa = K^\kappa{*}K^\kappa = K^\kappa{*}r{*}K^\kappa = (K^\kappa)^\tau = (K^\kappa * r)^\tau = (K^\kappa)^{\tau^\star} = (r * K^\kappa)^{\tau^\star} = (K' \cup K'')^\kappa$.

The four operations are *finitary*: they all take a finite number of arguments. Considering this, finite $\circ$-algebrae can be represented by a finite set and four multiplication tables. Hence, algorithms can input finite $\circ$-algebrae, unlike finite $\circ$-semigroups.

As for semigroups and $\circ$-semigroups, $\circ$-algebrae are given a notion of homomorphisms: a *homomorphism* between two $\circ$-algebrae S_1 and S_2 is a function that preserves the four operations (meaning that we have the equalities: $h(s{*}t) = h(s) * h(t)$, $h(s^\tau) = h(s)^\tau$, $h(s^{\tau^\star}) = h(s)^{\tau^\star}$, and $h(\{s_0, \ldots, s_{n-1}\}^\kappa) = \{h(s_0), \ldots, h(s_{n-1})\}^\kappa)$. A particular $\circ$-algebra is, for an alphabet $\mathbb{A}$, the set $\mathbb{A}^\circ$ with the operations $\cdot$, $(\cdot)^\omega$, $(\cdot)^{\omega^\star}$, and $(\cdot)^\eta$. Once again, a language $L \subseteq \mathbb{A}^\circ$ is *recognised* by a $\circ$-algebra S via a homomorphism $h\colon \mathbb{A}^\circ \to S$ if $L = h^{-1}(H)$ for some $H \subseteq S$.

Quite naturally, every $\circ$-semigroup S induces a $\circ$-algebra, with the four operations defined in the following way, where, in each case, the operation π is applied to a word over the alphabet S:

- for all $s, t \in S$, the product $s * t$ is defined as $\pi(s \cdot t)$;

– for every $s \in S$, s^τ and $s^{\tau^\star}$ are defined respectively as $\pi(s^\omega)$ and $\pi(s^{\omega^\star})$;
– for every finite $K \subseteq S$, K^κ is defined as $\pi(K^\eta)$.

We also say that the structure of $\circ$-semigroup (*i.e.* $\langle S, \pi\rangle$) *extends* the structure of $\circ$-algebra (*i.e.* $\langle S, *, \tau, \tau^\star, \kappa\rangle$). In the assumption that S is finite, Carton, Colcombet, and Puppis proved that this generalised product π is actually the unique generalised product that extends the $\circ$-algebra:

Theorem 1.27 ([CCP18, Theorem 24]). *Every finite $\circ$-algebra admits a unique extension into a $\circ$-semigroup.*

Considering that the algebraic objects we consider in the thesis are finite, we will not make the distinction between $\circ$-semigroups and $\circ$-algebrae: when one of the two is defined, we will use the generalised product π as much as the other four operations. A corollary of Theorem 1.27 is that the regularity of a language in $\mathbb{A}^\circ$ can be expressed by finite $\circ$-algebrae:

Corollary 1.28. *Let $L \subseteq \mathbb{A}^\circ$ be a language of countable words over an alphabet $\mathbb{A}$. Then L is regular if and only if it is recognised by some finite $\circ$-algebra.*

Corollary 1.28 implies that whenever we want to effectively represent a regular language of countable words $L \subseteq \mathbb{A}^\circ$, then it can be done by providing the structure of a finite $\circ$-algebra S, together with a representation of a homomorphism $h\colon \mathbb{A}^\circ \to S$ recognising L. The structure of S can be explicitly given, as written on page 48. The homomorphism h can be represented by providing the values for all single letters $\big(h(a)\big)_{a\in\mathbb{A}}$—such a representation uniquely determines h, as noted in a comment in the proof of Theorem 26 on page 22 of [CCP18]. Similarly, the respective set $H = h(L) \subseteq S$ can also be enumerated, since S is finite.

1.3.4 Ramsey's theorems for words

To conclude this section, we state two variants of Ramsey's theorem, in the context of finite semigroups and $\circ$-semigroups, which we will use later in the thesis.

Theorem 1.29 ([Sim84], see also [Lot97, Theorem 4.1.4]). *Let $\mathbb{A}$ be an alphabet, and h a homomorphism from $\mathbb{A}^+$ to some finite semigroup S. Then, for every natural number $n \geq 2$, there exist a natural number $N(n)$ such that for each word w over $\mathbb{A}$ of length at least $N(n)$, there exists an idempotent e in S and a decomposition $w = u \cdot w_0 \cdots w_{n-1} \cdot v$, where for all $i \in n$, w_i is non-empty and $h(w_i) = e$.*

Theorem 1.30 ([CPP08, Theorem 3.1]). *Let $\mathbb{A}$ be an alphabet, and h a homomorphism from $\mathbb{A}^\omega$ to some finite $\circ$-semigroup S. Then for each ω-word w over $\mathbb{A}$, there exists a decomposition $w = v \cdot w_0 \cdot w_1 \cdot w_2 \cdots$, where v and all the w_i's are non-empty finite words, such that $h(v) = s$ and $h(w_i) = e$ for all $i \in \mathbb{N}$, where s and e are two elements of S satisfying $s \cdot e = s$, $e \cdot e = e$, and $s \cdot e^\omega = \pi(w)$.*

1.4 Choice and uniformisation

In this section, we establish the main questions we aim to answer in the thesis. These questions are about the possibility of expressing choices in constructive ways. First, in Subsection 1.4.1, we explain which mathematical meaning we give to a choice, by presenting a quick history of the axiom of choice. We highlight the interest of being able to define choices in a constructive way, and we also define related notions, such as *uniformisations*. In Subsection 1.4.2, we reword this problem in terms of logic and languages of words. Finally, in Subsection 1.4.3, we recall a few known results about uniformisations and regular languages, which we use as a basis to our new results in the thesis.

1.4.1 Axiom of choice and uniformisations

A *choice function* over a set E is a function f from $\mathcal{P}(E) \setminus \{\emptyset\}$, the set of non-empty subsets of E, to E itself, such that for each non-empty $X \subseteq E$, $f(X) \in X$. In other terms, f distinguishes a particular element for each non-empty subset of E.

To the basic Zermelo-Fraenkel set theory is often added what we call the axiom of choice, introduced for the first time by Zermelo in 1904. It states that choice functions always exist:

Axiom 1.31 (*Choice*, [Zer04]). *Every set admits a choice function.*

Historically, the notion of choice functions used by Zermelo was not the same. He described a choice function over a family $(A_i)_{i \in I}$ of non-empty sets as a function from I to $\bigcup_{i \in I} A_i$, the union of all the A_i's, such that $f(i) \in A_i$ for each $i \in I$. However, one can quickly verify that the two versions of the axiom are equivalent, by defining $I := \mathcal{P}(E) \setminus \{\emptyset\}$ and $A_X = X$ for one direction of the equivalence, and by defining $E := \bigcup_{i \in I} A_i$ for the other. The reason we use this different notion of choice function is that it fits better our purpose in the thesis. It is not a problem, not only because we saw that these two notions lead to equivalent axioms, but also because the axiom of choice has always been known under many equivalent statements.

Now, we state some other statements which are well-known to be equivalent to the axiom of choice in **ZF**. The first one is known as Zermelo's theorem, proved in [Zer04] to be an implication of the axiom:

Theorem 1.32 (*Zermelo*, [Zer04])**.** *Every set admits a well order.*

On the other hand, the fact that Zermelo's theorem implies the axiom of choice is clear, for if a set E admits a well order, then the function that maps each non-empty subset of E to its least element in this order is a choice function over E. Therefore, the axiom of choice and Zermelo's theorem are two equivalent statements in **ZF**.

Another example of a result equivalent to the axiom of choice is the Kuratowski-Zorn lemma:

Lemma 1.33 (*Kuratowski-Zorn*, [Kur22])**.** *Let ϑ be a non-empty partially ordered set. We suppose that every* chain *of ϑ admits an upper bound in ϑ. Then ϑ admits at least one maximal element.*

In this lemma, a *chain* C of ϑ is a sequence $(x_j)_{j\in J}$ indexed by a linear order J (not necessarily countable) that is *increasing*: $x_j < x_k$ for all $j < k$ in J, and an *upper bound* of C in ϑ is an element $m \in E$ such that $x_j \leq m$ for each $j \in J$. The Kuratowski-Zorn lemma is often simply called Zorn's lemma, as Zorn gave in [Zor35] many applications of this result—stated as the *maximum principle*. However, Kuratowski was the first to give in 1922 a proof that this result is implied by Zermelo's theorem (and therefore by the axiom of choice). A standard proof that in **ZF** the Kuratowski-Zorn lemma implies the axiom of choice is more straightforward:

Proof. Let us suppose that the Kuratowski-Zorn lemma is true, and let $(A_i)_{i\in I}$ be any family of non-empty sets. We show that it admits a choice function in Zermelo's sense.

We define ϑ as the set of partial functions f from I to $\bigcup_{i\in I} A_i$ such that for all $i \in \mathrm{Dom}(f)$, $f(i) \in A_i$. We provide ϑ with the partial order defined by $f \leq g$ if $\mathrm{Dom}(f) \subseteq \mathrm{Dom}(g)$ and for all $i \in \mathrm{Dom}(f)$, $f(i) = g(i)$. The set ϑ is non-empty, and each chain $(f_j)_{j\in J}$ of ϑ admits an upper bound $f \in \vartheta$, defined by $\mathrm{Dom}(f) = \bigcup_{j\in J} \mathrm{Dom}(f_j)$, and by $f(i) = f_j(i)$ if $i \in \mathrm{Dom}(f_j)$. Therefore, because the Kuratowski-Zorn lemma is true by assumption, ϑ must admit a maximal element f, which is necessarily defined over the whole set I. Indeed, if k was an element of $I \setminus \mathrm{Dom}(f)$, then we could consider any $x \in A_k$, and we would obtain $f < f'$, with $f' \in \vartheta$ defined by $\mathrm{Dom}(f') = \mathrm{Dom}(f) \sqcup \{k\}$, $f(k) = x$, and $f'(i) = f(i)$ for $i \in \mathrm{Dom}(f)$. This proves that f is a choice function over $(A_i)_{i\in I}$, and we can conclude the proof. □

An important result is that, if we assume that the foundation **ZF** is *consistent* (*i.e.* no contradiction can be deduced from it), then the axiom of choice is *independent* of **ZF**, meaning that it can neither be proven nor be refuted in this theory:

Theorem 1.34 ([Göd38], [Coh63])**.** *If* **ZF** *is consistent, then the axiom of choice is independent of* **ZF**.

Gödel was the first to prove, in 1938, that, under the assumption of the consistency of **ZF**, the addition of the axiom of choice in the theory would not lead to a contradiction, and Cohen proved 25 years later that, under the same assumption, the axiom of choice could not be proven in **ZF**, which completes the theorem.

Hence, accepting the axiom of choice or not is rather a philosophical question than a mathematical one, and nowadays, it is considered as a valid principle for most mathematicians. Nevertheless, as we stated it in the introduction, the axiom of choice can be used to justify the existence of mathematical objects that are not perceptible to us and are uneasy to manipulate. As examples of such objects, we can cite Vitali sets: intuitively, these are subsets of $\mathbb{R}$, the set of real numbers, that cannot be given a consistent measure, see [Leb02] for the definition of the Lebesgue measure, and [Vit05] for the construction of the first Vitali set.

This difficulty justifies the search of more concrete choice functions, whose existence can be proved without reference to the axiom. Our aim in the thesis is to find conditions under which such choice functions can be explicitly constructed:

Problem 1.35. *Under which condition can we constructively provide choice functions?*

In this thesis, as we are going to state it more formally in Subsection 1.4.2, we will focus on the case of words and linear orders, and our aim will be to describe choice functions with regular formulae, when it is possible. In this quest, we will also focus on another object, called a *uniformisation*.

Let $R \subseteq E \times F$ be a binary relation between two sets E and F. We define $\Pi_E(R)$, the *projection* of R onto E, as the set $\{x \in E \mid \text{there exists some } y \in Y \text{ such that } \langle x, y\rangle \in R\}$, and, if $x \in E$, each $y \in F$ such that $\langle x, y\rangle \in R$ is called a *candidate* for x in R. We say that a function f from $\Pi_E(R)$ is a *uniformisation* of R, or that it *uniformises* R, if $\langle x, f(x)\rangle \in R$ for each $x \in \Pi_E(R)$ (see Figure 1.2). In other words, f selects a unique candidate in R for each element of its projection $\Pi_E(R)$. This is why in the literature, f is sometimes also called a *selector* of R.

Proposition 1.36. *The axiom of choice is true if and only if every binary relation admits a uniformisation.*

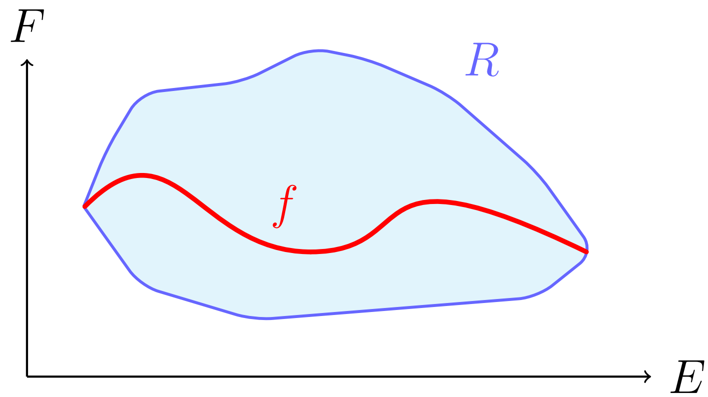


Figure 1.2: The function f uniformises the relation R.

Proof. Let us suppose that the axiom of choice is true, and let $R \subseteq E \times F$, with E and F being two sets. We consider the family $(\{y \in F \mid \langle x, y\rangle \in R\})_{x \in \Pi_E(R)}$, and then a choice function in Zermelo's sense over this family is in fact a uniformisation of R.

To prove that the existence of uniformisations implies the axiom of choice (in our sense), it is enough to consider the binary relation $R := \big\{\langle X, x\rangle \mid X \subseteq E \text{ and } x \in X\big\} \subseteq \mathcal{P}(E) \times E$: a uniformisation over R is in fact a choice function over E. $\square$

Proposition 1.36 justifies the following reformulation of Problem 1.35, yet not less elusive:

Problem 1.37 (*Quest of uniformisations*)**.** *In which condition can we constructively provide uniformisations?*

As stated above, we will focus on the field of linear orders and words. In the next subsection, we give some proper instance of Problem 1.37 and explicit how we understand this notion of "constructibility" in said field.

1.4.2 The case of words and regular languages

In this subsection, we explain how Problems 1.35 and 1.37 can be understood when working with linear orders and words. In particular, we explain what it means for a binary relation between sets of words to be definable in some formalism.

Let Σ be a signature, and $\mathcal{M}$ a model of Σ. A choice function f over $\mathcal{M}$ is *regular* if there exists a formula $\varphi^{\mathcal{M}}_{\text{choice}}(X, x)$ with exactly two free variables, one second-order and one first-order, such that for every non-empty $X \subseteq \mathcal{M}$, there exists a unique $x \in \mathcal{M}$ such that $\mathcal{M} \vDash \varphi^{\mathcal{M}}_{\text{choice}}(X, x)$, and moreover this particular x is $f(X)$, the image of X under f (in particular it must belong to X). We can also say that f is an MSO[Σ] choice function, if we want to emphasise the signature. More generally, we say that f is a C choice function if said formula $\varphi^{\mathcal{M}}_{\text{choice}}(X,x)$ belongs to C, C being a class of MSO formulae. For instance, if $\varphi^{\mathcal{M}}_{\text{choice}}$ does not contain any second-order quantifier, *i.e.* the only second-order variable in $\varphi^{\mathcal{M}}_{\text{choice}}$ is the free variable X, then f is called *a first-order choice function* (or FO[Σ] choice function, FO choice function).

Example 1.38. *Let ϖ be a well order. By definition, every non-empty subset X of ϖ admits a least element. Therefore, the formula $\varphi(X,x) := x \in X \wedge \forall y{\in}X.\ x \leq y$ is an* FO[<] *choice function over ϖ.*

Now, we can give an example of how Problem 1.35 can be concretely rephrased: "considering a linear order λ, on which condition does there exist a choice function over λ definable in a given formalism C?". In particular, Example 1.38 tells us that if said order is a well order, then it admits a regular choice function (and even an FO one). We will try to find other conditions allowing such functions.

To give a more explicit meaning to Problem 1.37, we also provide a notion of a *regular uniformisation*, in the context of words.

Let $\mathbb{A}$ and $\mathbb{B}$ be two alphabets, and let λ be a linear order. To each pair of words $\langle w, \sigma\rangle$ in $\mathbb{A}^\lambda \times \mathbb{B}^\lambda$ corresponds a unique λ-word p over the product alphabet $\mathbb{A}\times\mathbb{B}$, defined by $p(x) = \langle w(x), \sigma(x)\rangle$ for each $x \in \lambda$, and reciprocally, to each λ-word over the product alphabet $\mathbb{A}\times\mathbb{B}$ corresponds a unique pair of words in $\mathbb{A}^\lambda \times \mathbb{B}^\lambda$. Therefore, it makes sense to identify this pair $\langle w, \sigma\rangle$ with this word p. We introduce convenient notations: we now write $\binom{\mathbb{A}}{\mathbb{B}}$ for the product alphabet $\mathbb{A} \times \mathbb{B}$; if $a \in \mathbb{A}$ and $b \in \mathbb{B}$, then we write $\binom{a}{b}$ for the corresponding letter in $\binom{\mathbb{A}}{\mathbb{B}}$; and if $w \in \mathbb{A}^\lambda, \sigma \in \mathbb{B}^\lambda$, then we denote by $\binom{w}{\sigma}$ the λ-word over $\binom{\mathbb{A}}{\mathbb{B}}$ corresponding to the pair $\langle w, \sigma\rangle$. We continue by identifying every binary relation $R \subseteq \mathbb{A}^\lambda \times \mathbb{B}^\lambda$ with the λ-language $\left\{\binom{w}{\sigma} \in \binom{\mathbb{A}}{\mathbb{B}}^\lambda \mid wR\sigma\right\}$. Similarly, a language $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ is seen as a relation of finite words whose property is that only pairs of words of same length are in this relation. Now, we can say that a relation between words is *regular* (resp. in **FO**[<], **FO**[s]...) if so is the corresponding language.

We adapt to languages of words our notion of uniformisation. Let $\mathbb{A}$ and $\mathbb{B}$ be two alphabets, and let F, R be two relations included in $\binom{\mathbb{A}}{\mathbb{B}}^\lambda$. We say that F is a *uniformisation* of R, or that it *uniformises* R, if the two following conditions are fulfilled: $F \subseteq R$; for every w in $\mathbb{A}^\lambda$, there exists some $\sigma \in \mathbb{B}^\lambda$ such that $\binom{w}{\sigma} \in R$ if and only if there exists a unique $\sigma \in \mathbb{B}^\lambda$ such that $\binom{w}{\sigma} \in F$. The second condition is in fact the conjunction of the following two others: first, the projections $\Pi_{\mathbb{A}^\lambda}(F) := \{w \in \mathbb{A}^\lambda \mid$ there exists some σ in $\mathbb{A}^\lambda$ such that $\binom{w}{\sigma} \in$

$F\}$ and $\Pi_{\mathbb{A}^\lambda}(R)$ must be the same, and, second, F must be *functional*, *i.e.* it cannot relate a word $w \in \Pi_{\mathbb{A}^\lambda}(F)$ with two distinct words σ in $\mathbb{B}^\lambda$ or more. Part of the literature might call *partially functional* a relation satisfying this last condition, and rather reserve the adjective *functional* for relations whose projections are moreover the full set $\mathbb{A}^\lambda$. In the thesis, we will never take into account whether the projection of a relation is the full set or not, therefore, we keep our definition of functional. Once again, we consider an analogous definition of a uniformisation for a binary relations $R \subseteq \left(\begin{smallmatrix}\mathbb{A}\\ \mathbb{B}\end{smallmatrix}\right)^+$ of finite words.

Now, if $\mathbf{C}_1$ and $\mathbf{C}_2$ are two classes of languages, we say that $\mathbf{C}_2$ *uniformises* $\mathbf{C}_1$ if every relation in $\mathbf{C}_1$ (in the sense that it corresponds to a language in $\mathbf{C}_1$ over some product alphabet) admits a uniformisation in $\mathbf{C}_2$. We also say that $\mathbf{C}_1$ is (or can be) uniformised in $\mathbf{C}_2$. This definition allows us to state the following problem:

Problem 1.39. *Given $\mathbf{C}_1$ and $\mathbf{C}_2$ two classes of languages, can $\mathbf{C}_1$ be uniformised in $\mathbf{C}_2$?*

In the thesis, we will mainly study the question when $\mathbf{C}_1$ and $\mathbf{C}_2$ are classes of regular languages, namely when $\mathbf{C}_1$ and $\mathbf{C}_2$ are $\mathbf{MSO}[<]$, $\mathbf{FO}[<]$, $\mathbf{FO}[\mathsf{s}]$... The following fact gives us a natural restriction when $\mathbf{C}_1$ and $\mathbf{C}_2$ are varieties and are closed under extensions of alphabets (see page 40 for the definition):

Fact 1.40. *Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two varieties of languages closed under extensions of alphabets and such that $\mathbf{C}_2$ uniformises $\mathbf{C}_1$. Then $\mathbf{C}_1 \subseteq \mathbf{C}_2$.*

Proof. Let $L \in \mathbf{C}_1$, over the alphabet $\mathbb{A}$. We define L_{identity} as the relation containing all the pairs $\left(\begin{smallmatrix}w\\ w\end{smallmatrix}\right)$, for $w \in L$. This language L_{identity}, over the alphabet $\left\{\left(\begin{smallmatrix}a\\ a\end{smallmatrix}\right) \mid a \in \mathbb{A}\right\}^+$, is the preimage of L under the homomorphism induced from the function $\left(\begin{smallmatrix}a\\ a\end{smallmatrix}\right) \mapsto a$. Since $\mathbf{C}_1$ is a variety of languages, L_{identity} is in $\mathbf{C}_1$. Since $\mathbf{C}_1$ is closed under extensions of alphabets, the language $\left\langle L_{\text{identity}}, \left(\begin{smallmatrix}\mathbb{A}\\ \mathbb{A}\end{smallmatrix}\right)\right\rangle$ is also in $\mathbf{C}_1$, and, by the assumption, it must admit a uniformisation $F \in \mathbf{C}_2$. It is clear that F must be L_{identity} itself: every word $w \in \Pi_{\mathbb{A}^+}(L_{\text{identity}}) = \mathbb{A}^+$ is actually the only σ such that $\left(\begin{smallmatrix}w\\ \sigma\end{smallmatrix}\right) \in L_{\text{identity}}$. Hence, $\left\langle L_{\text{identity}}, \left(\begin{smallmatrix}\mathbb{A}\\ \mathbb{A}\end{smallmatrix}\right)\right\rangle \in \mathbf{C}_2$ and, therefore, also L is in $\mathbf{C}_2$, since it is the preimage of $\left\langle L_{\text{identity}}, \left(\begin{smallmatrix}\mathbb{A}\\ \mathbb{A}\end{smallmatrix}\right)\right\rangle \in \mathbf{C}_2$ by the homomorphism from $\mathbb{A}^+$ to $\left(\begin{smallmatrix}\mathbb{A}\\ \mathbb{A}\end{smallmatrix}\right)^+$ induced from the function $a \mapsto \left(\begin{smallmatrix}a\\ a\end{smallmatrix}\right)$, and this concludes the proof. $\square$

Notice that the assumption that the class $\mathbf{C}$ is closed under extensions of alphabets does not follow from the fact that $\mathbf{C}$ is a variety of languages. Indeed, the class $\mathbf{EF}$ consisting of the empty languages $\langle\emptyset, \mathbb{A}\rangle$ and full languages $\langle\mathbb{A}^+, \mathbb{A}\rangle$ over all alphabets $\mathbb{A}$ is an example of a variety that it is not closed under extensions of alphabets.

As the totality of classes of languages considered in this thesis are varieties of languages and are closed under extensions of alphabets, Problem 1.39 is only relevant when $\mathbf{C}_1$ is included in $\mathbf{C}_2$. For instance, over finite words, it does make sense to ask whether $\mathbf{FO}^2[\mathsf{s}]$ is uniformised by $\mathbf{FO}[<]$, when the reciprocal question is immediately answered negatively. When $\mathbf{C}_1$ cannot be uniformised in $\mathbf{C}_2$, a less restrictive version of the problem can be asked:

Problem 1.41. *Given two classes of languages* $\mathbf{C}_1$ *and* $\mathbf{C}_2$*, can we characterise the relations in* $\mathbf{C}_1$ *that admit a uniformisation in* $\mathbf{C}_2$*?*

Now, the question is relevant not only when $\mathbf{C}_1 \nsubseteq \mathbf{C}_2$. As an example, it could be interesting to know which individual regular relations admit a uniformisation in $\mathbf{FO}[\mathsf{s}]$.

Naturally, we can also focus on Problems 1.39 and 1.41 in the special case $\mathbf{C}_1 = \mathbf{C}_2$:

Problem 1.42. *Given* $\mathbf{C}$ *a class of languages, does* $\mathbf{C}$ *uniformise itself?*

If a class uniformises itself, we also say that it satisfies the *uniformisation property*, or that it is *self-uniformisable*.

Problem 1.43. *Given* $\mathbf{C}$ *a class of languages, can we characterise the relations in* $\mathbf{C}$ *that admit a uniformisation also in* $\mathbf{C}$*?*

1.4.3 Known results about regular uniformisations

In this subsection, we state some already known results about uniformisations. Namely, that $\mathbf{MSO}[<]$, the class of regular languages, satisfies the uniformisation property over both finite words and ω-words. We say that finite and ω-words satisfy the *regular-uniformisation property*. A proof of the result for ω-words can be found in [Rab07], but there is no reliable source for the former case, over finite words, since it is mostly considered as folklore. For the sake of completeness, we propose a proof of this result in this section.

Let $\mathbb{A}$, $\mathbb{B}$ be two alphabets, and let λ be a linear order. Without loss of generality, we assume that $\mathbb{B} = \{b_0, \ldots, b_{k-1}\}$, with $k = |\mathbb{B}|$. If $(X_i)_{i \in k}$ is a family of subsets of λ that forms a *k-partition* of λ (or simply a *partition* of λ), meaning if the X_i's are pairwise disjoint and $\bigsqcup_{i \in k} X_i = \lambda$, then it induces a λ-word over $\mathbb{B}$, which we write $\sigma(X_0, \ldots, X_{k-1})$, in which each position is labelled by b_i if and only if it is in X_i. Reciprocally, to any λ-word σ over $\mathbb{B}$ corresponds a partition $(X_i)_{i \in k}$, defined by $X_i = \{x \in \lambda \mid \sigma(x) = b_i\}$. A first thing to notice is that forming a k-partition can be defined regularly.

Fact 1.44. *Being a k-partition is regular.*

Proof. Let $X_0, \ldots, X_{k-1}$ be k second-order variables. The fact that their interpretations in λ form a k-partition is defined by the formula $\forall x.\ \bigvee_{i \in k} x \in X_i \wedge \bigwedge_{j \in k, j \neq i} x \notin X_j$, which we denote by $\varphi^k_{\text{part}}(X_0, \ldots, X_{k-1})$. $\square$

Now, we can use this correspondence between words over $\mathbb{B}$ and k-partitions to express regular properties on a third λ-word, as formalised in the following two claims:

Claim 1.45. *Let $(X_i)_{i\in k}$ be a k-partition of λ, and φ be a sentence in* $\mathrm{MSO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$. *Then there exists a formula $\varphi'(X_0, \ldots, X_{k-1})$ in the same class such that for all $\binom{w}{\sigma} \in \binom{\mathbb{A}}{\mathbb{B}}^\lambda$, we have $\binom{w}{\sigma} \vDash \varphi'(X_0, \ldots, X_{k-1})$ if and only if $\binom{w}{\sigma(X_0,\ldots,X_{k-1})} \vDash \varphi$.*

Proof. The formula $\varphi'(X_0, \ldots, X_{k-1})$ is obtained from φ by replacing each subformula of the shape $\binom{a}{b_i}(x)$ by the formula $\bigvee_{b\in\mathbb{B}} \binom{a}{b}(x) \wedge x \in X_i$. □

Claim 1.46. *Let $(X_i)_{i\in k}$ be a k-partition of λ, and φ be a formula in* $\mathrm{MSO}[\binom{\mathbb{B}}{\mathbb{B}}, <]$. *Then there exists a formula $\varphi''(X_0, \ldots, X_{k-1})$ in* $\mathrm{MSO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$ *such that for all $\binom{w}{\sigma} \in \binom{\mathbb{A}}{\mathbb{B}}^\lambda$, we have $\binom{w}{\sigma} \vDash \varphi''(X_0, \ldots, X_{k-1})$ if and only if $\binom{\sigma}{\sigma(X_0,\ldots,X_{k-1})} \vDash \varphi$.*

Proof. This time, the formula $\varphi''(X_0, \ldots, X_{k-1})$ is obtained from φ by replacing each subformula of the shape $\binom{b}{b_i}(x)$, by the formula $\bigvee_{a\in\mathbb{A}} \binom{a}{b}(x) \wedge x \in X_i$. □

Before finally proving the theorem, we use these k-partitions to show that **MSO**[<] is closed under projections, a result which we will use later in the thesis:

Proposition 1.47. *Let φ be a formula in* $\mathrm{MSO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$, *$\mathbb{A}$ and $\mathbb{B}$ being two alphabets. Then there exists an* $\mathrm{MSO}[\mathbb{A}, <]$ *formula φ_{proj} such that for all $w \in \mathbb{A}^\lambda$, $w \vDash \varphi_{\mathrm{proj}}$ if and only if $\binom{w}{\sigma} \vDash \varphi$ for some $\sigma \in \mathbb{B}^\lambda$.*

Proof. Here again we assume that $\mathbb{B} = \{b_0, \ldots, b_{k-1}\}$, with $k = |\mathbb{B}|$. If we replace each subformula of φ of the shape $\binom{a}{b_i}(x)$ by $a(x) \wedge x \in X_i$, then we obtain a formula $\varphi'''(X_0, \ldots, X_{k-1})$ such that, if $X_0, \ldots, X_{k-1}$ is a k-partition of λ, then $w \vDash \varphi'''(X_0, \ldots, X_{k-1})$ if and only if $\binom{w}{\sigma(X_0,\ldots,X_{k-1})} \vDash \varphi$. Then, it suffices to define φ_{proj} as $\exists X_0, \ldots, X_{k-1}.\ \varphi^k_{\mathrm{part}}(X_0, \ldots, X_{k-1}) \wedge \varphi'''(X_0, \ldots, X_{k-1})$. □

We can now prove the first theorem of this subsection:

Theorem 1.48. *Over finite words,* $\mathrm{MSO}[<]$ *satisfies the uniformisation property: for any alphabets $\mathbb{A}$, $\mathbb{B}$, every regular relation $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ admits a regular uniformisation $F \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$.*

Proof. The proof is based on the fact that for all $n \in \omega$, if we consider $\mathbb{B}$ with a linear order, then $\mathbb{B}^n$ with $\leq_{\mathrm{lex}}$ is a well order (see page 25). If $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ is regular, then for each $w \in \Pi_{\mathbb{A}^+}(R)$, there exists a word $\sigma \in \mathbb{B}^{|w|}$ such that $\binom{w}{\sigma} \in R$ and for all $\sigma' \in \mathbb{B}^{|w|}$ satisfying $\binom{w}{\sigma'} \in R$, we have $\sigma \leq_{\mathrm{lex}} \sigma'$. Our aim is to define this condition in $\mathrm{MSO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$.

Let $k = |\mathbb{B}|$, and let us assume that $\mathbb{B}$ is $\{b_0, \dots, b_{k-1}\}$, without loss of generality. As written before, we identify words over $\mathbb{B}$ with k-partitions. Therefore, we want to express regularly that for each k-partition $(X_i)_{i \in k}$, if $\binom{w}{\sigma(X_0,\dots,X_{k-1})} \in R$, then $\sigma \leq_{\text{lex}} \sigma(X_0, \dots, X_{k-1})$.

First, quantifying over k-partitions is possible, as it is stated in Fact 1.44, and, if $\varphi \in \mathrm{MSO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$ defines R, then, using Claim 1.45, we can construct $\varphi'(X_0, \dots, X_{k-1})$ defining the condition $\binom{w}{\sigma(X_0,\dots,X_{k-1})} \in R$.

It remains to deal with the lexicographic order. We consider that $\mathbb{B} = \{b_0, b_1, \dots, b_{k-1}\}$ is ordered by $b_i < b_j$ when $i < j$. First, make it clear that an $\mathrm{MSO}[\binom{\mathbb{B}}{\mathbb{B}}, <]$ formula φ_{lex} is satisfied by $\binom{\sigma}{\sigma'}$ if and only if $\sigma \leq_{\text{lex}} \sigma'$ can be constructed. Indeed, for a single position x, conditions $\sigma(x) = \sigma'(x)$ and $\sigma(x) < \sigma'(x)$ can both be defined by MSO formulae, respectively by $\varphi_=(x) := \bigvee_{i \in k} \binom{b_i}{b_i}(x)$ and by $\varphi_<(x) := \bigvee_{i \in j \in k} \binom{b_i}{b_j}(x)$. Therefore, the formula $\varphi_{\text{lex}} := \forall x.\ \varphi_=(x) \vee \exists y.\ \varphi_<(y) \wedge \forall x{<}y.\ \varphi_=(x)$ fulfils our need (in fact, it is even a first-order formula). Now, with Claim 1.46, we obtain a formula $\varphi''_{\text{lex}}(X_0, \dots, X_{k-1})$ such that $\binom{w}{\sigma} \models$ if and only if $\sigma \leq_{\text{lex}} \sigma'(X_0, \dots, X_{k-1})$.

Putting it all together, we obtain a formula φ_{unif}, defined as:

$$\begin{aligned} &\varphi \wedge \forall X_0, \dots, X_{k-1}.\ \varphi^k_{\text{part}}(X_0, \dots, X_{k-1}) \\ &\qquad\qquad\qquad \Longrightarrow \big(\varphi'(X_0, \dots, X_{k-1}) \Longrightarrow \varphi''_{\text{lex}}(X_0, \dots, X_{k-1})\big), \end{aligned}$$

that selects for each $w \in \Pi_{\mathbb{A}^+}(R)$ the lexicographically least candidate in R: $\mathcal{L}(\varphi_{\text{unif}})$ uniformises $R = \mathcal{L}(\varphi)$, and the proposition is proven. □

Finally, we write the result over ω-words. This result was first stated in [LS98], claiming that it came from [BL67]. However, this reference was not legitimate, since [BL67] does not involve uniformisations, but another non-equivalent problem, about *games with finite-memory strategies*. This is the reason why Rabinovich provided a second proof of the theorem in the Appendix of [Rab07].

Theorem 1.49 ([Rab07, Theorem 27]). *Over ω-words,* MSO *satisfies the uniformisation property: for any alphabets $\mathbb{A}$, $\mathbb{B}$, every regular relation $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^\omega$ admits a regular uniformisation $F \subseteq \binom{\mathbb{A}}{\mathbb{B}}^\omega$.*

There is another field of research in which one is interested in providing uniformisations to *rational* relations of words, meaning relations given by *asynchronous transducers* (see for instance [DFKL20] for a definition). However, since we do not consider this kind of relations in the thesis, we do not get into the details of these. We simply state that, in that branch, results similar to Theorems 1.48 and 1.49 have been proven: [Kob69] proves that rational relations of finite words admit rational uniformisations, while [CG99, Theorem 5] proves that rational relations of ω-words admit rational uniformisations.

Chapter 2

Uniformising in fragments of First-Order Logic

In this chapter, we give some results, both positive and negative, when it comes to uniformise regular relations of finite words in First-Order Logic, $\mathbf{FO}[<]$, and some of its fragments, defined in Subsection 1.2.4. In particular, we give answers to Problems 1.39 and 1.41 in some particular cases.

First, we prove the following three negative results, originally published in [Mic18]:

Proposition 2.1. *Over finite words,* $\mathbf{FO}[<]$ *does not uniformise* $\mathbf{FO}^2[\mathsf{s}]$.

Proposition 2.2. *Over finite words,* $\mathbf{FO}[\mathsf{s}]$ *does not uniformise* $\mathbf{FO}^2[\,]$.

Proposition 2.3. *Over finite words,* $\mathbf{FO}^2[<,\mathsf{s}]$ *does not uniformise* $\mathbf{FO}^2[<] \cap \mathbf{FO}^2[\mathsf{s}]$.

These three propositions are depicted in Figure 2.1, where a crossed arrow from a class $\mathbf{C}_1$ to a class $\mathbf{C}_2$ represents the impossibility to uniformise in $\mathbf{C}_2$ all the relations of $\mathbf{C}_1$. On the opposite, a green arrow from a class $\mathbf{C}_1$ to a class $\mathbf{C}_2$ represents the fact that all relations from $\mathbf{C}_1$ admit uniformisations in $\mathbf{C}_2$, see Proposition 2.5 below.

As a consequence of these propositions, none of the fragments of First-Order Logic introduced in Subsection 1.2.4 admit the uniformisation property, since they are all between the two classes involved in some of these three negative results.

Indeed, if $\mathbf{C}$ satisfies the uniformisation property, then it is immediate that for all classes $\mathbf{C}_1$ and $\mathbf{C}_2$ such that $\mathbf{C}_1 \subseteq \mathbf{C} \subseteq \mathbf{C}_2$, $\mathbf{C}_2$ does uniformise $\mathbf{C}_1$: if R is a relation in $\mathbf{C}_1$, then it is also in $\mathbf{C}$ and therefore admits a uniformisation $F \in \mathbf{C}$, which is in $\mathbf{C}_2$.

Corollary 2.4. *Over finite words, none of the following classes satisfy the uniformisation property:* $\mathbf{FO}[<]$, $\mathbf{FO}[\mathsf{s}]$, $\mathbf{FO}[\,]$, $\mathbf{FO}^k[<]$, $\mathbf{FO}^k[\mathsf{s}]$, $\mathbf{FO}^k[<,\mathsf{s}]$, $\mathbf{FO}^k[<] \cap \mathbf{FO}^k[\mathsf{s}]$, *and* $\mathbf{FO}^k[\,]$, *for* $k \geq 2$.

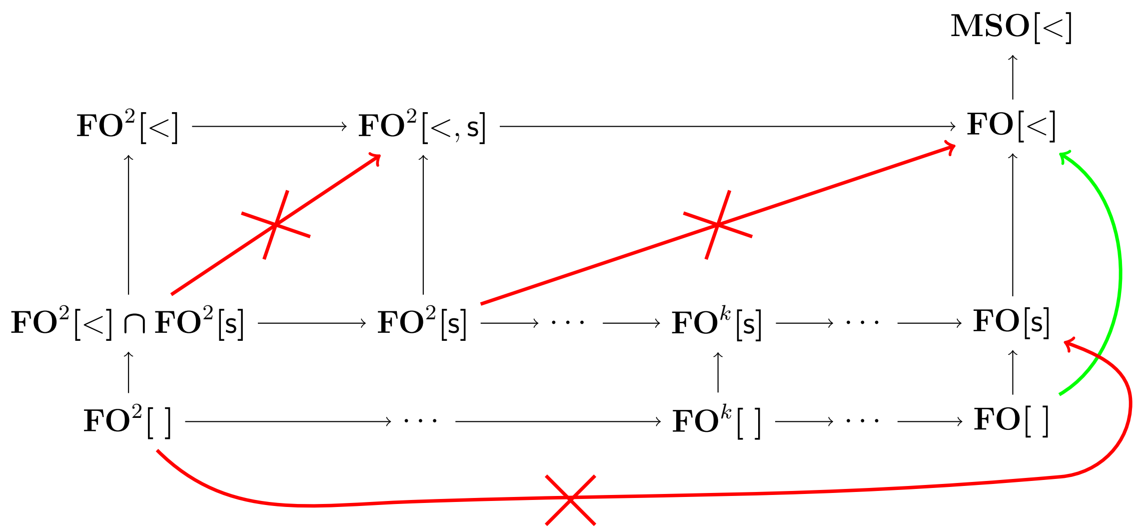


Figure 2.1: Illustration of Propositions 2.1 to 2.3, and 2.5.

This corollary contrasts with Theorem 1.48, stating that, over finite words, the class **MSO**$[<]$ *does* satisfy the uniformisation property. Therefore, this chapter helps us to understand what is missing in First-Order Logic in order to provide witnesses.

On the other hand, we prove a positive result, on the possibility to uniformise every relation of **FO**$[\,]$ in the class **FO**$[<]$:

Proposition 2.5. *Over finite words, the class* **FO**$[<]$ *uniformises the class* **FO**$[\,]$.

And, finally, we prove another positive result. It states the decidability of uniformising regular relations in the former class **FO**$[\,]$:

Theorem 2.6. *Let R be a regular relation of finite words. It is decidable whether R admits a uniformisation in* **FO**$[\,]$.

Proposition 2.5 and Theorem 2.6 have not been published yet.

Finally, we discuss the possibility to extend this decidability result to other fragments of First-Order Logic.

This chapter is divided into four sections. Section 2.1 is dedicated to the proof of the three negative propositions. In Section 2.2, we characterise the languages in the class **FO**$[\,]$ in terms of functions which we call $\langle \mathbb{A}, N\rangle$*-maps*, and we show some basic properties of these maps. Then, we use this characterisation in Sections 2.3 and 2.4, where we prove Proposition 2.5 and Theorem 2.6 respectively.

2.1 Negative results

In this first section, we prove Propositions 2.1, 2.2, and 2.3. It is divided into three subsections, each one dedicated to each of these negative results.

The strategy in these three proofs is similar: to show that a class $\mathbf{C}_2$ does not uniformise a class $\mathbf{C}_1$, we exhibit some alphabets $\mathbb{A}$ and $\mathbb{B}$, a particular relation R in $\mathbf{C}_1$ over the product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$, and we show that it cannot admit any uniformisation in $\mathbf{C}_2$, typically using algebraic arguments stated in Section 1.3.

2.1.1 The class $\mathbf{FO}[<]$ does not uniformise the class $\mathbf{FO}[\mathsf{s}]$

First, we give a proof of Proposition 2.1, the first negative result of this chapter. As stated above, we prove that a particular relation $R \in \mathbf{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, \mathsf{s}]$ does not admit any uniformisation in $\mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$.

Consider the alphabets $\mathbb{A} = \{a, b\}$ and $\mathbb{B} = \{0, 1, \rhd, \lhd\}$. We define the relation R_1 as the regular language

$$\left(\binom{a}{0} \cdot \binom{a}{1}\right)^+ \cdot \binom{b}{\lhd}^+,$$

over the product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$, and the relation R_2 as the symmetric language

$$\binom{a}{\rhd}^+ \cdot \left(\binom{b}{0} \cdot \binom{b}{1}\right)^+ .$$

Less formally, R_1 counts an even and positive number of a's before any positive number of b's, while, on the opposite R_2 counts an even and positive number of b's after any positive number of a's. Notice that these two relations are disjoint. Then, we define R, the relation which interests us, as the union $R_1 \sqcup R_2$ of these two languages.

A generic description of elements of R is drawn in Figure 2.2, with the left pair being a typical element of R_1 and the right pair a typical element of R_2.

w_1:	a	a	$\cdots$	a	a	b	b	$\cdots$	b	b
σ_1:	0	1	$\cdots$	0	1	$\lhd$	$\lhd$	$\cdots$	$\lhd$	$\lhd$

w_2:	a	a	$\cdots$	a	a	b	b	$\cdots$	b	b
σ_2:	$\rhd$	$\rhd$	$\cdots$	$\rhd$	$\rhd$	0	1	$\cdots$	0	1

Figure 2.2: Generic pairs in the relation R.

Since two variables and the successor function are enough to express the alternations of 0's and 1's, this relation is definable in $\mathrm{FO}^2[\mathsf{s}]$:

Fact 2.7. *The relation R is in* $\mathbf{FO}^2[\left(\begin{smallmatrix}\mathbb{A}\\\mathbb{B}\end{smallmatrix}\right), \mathsf{s}]$.

Proof. The relation R_1 is defined by the following $\mathrm{FO}^2[\left(\begin{smallmatrix}\mathbb{A}\\\mathbb{B}\end{smallmatrix}\right), \mathsf{s}]$ formula:

$$\begin{aligned}
\varphi_1 :=& \\
&\quad \exists x.\ \left(\begin{smallmatrix}a\\0\end{smallmatrix}\right)(x) \wedge \forall y.\ \neg \mathsf{s}(y,x) \\
&\wedge\ \forall x.\ \left(\begin{smallmatrix}a\\0\end{smallmatrix}\right)(x) \implies \exists y.\ \mathsf{s}(x,y) \wedge \left(\begin{smallmatrix}a\\1\end{smallmatrix}\right)(y) \\
&\wedge\ \forall x.\ \left(\begin{smallmatrix}a\\1\end{smallmatrix}\right)(x) \implies \exists y.\ \mathsf{s}(x,y) \wedge \Big(\left(\begin{smallmatrix}a\\0\end{smallmatrix}\right)(y) \vee \left(\begin{smallmatrix}b\\\lhd\end{smallmatrix}\right)(y)\Big) \\
&\wedge\ \forall x,y.\ \Big(\left(\begin{smallmatrix}b\\\lhd\end{smallmatrix}\right)(x) \wedge \mathsf{s}(x,y)\Big) \implies \left(\begin{smallmatrix}b\\\lhd\end{smallmatrix}\right)(y).
\end{aligned}$$

A symmetric formula φ_2 defines R_2, and, finally, the disjunction $\varphi_1 \vee \varphi_2$ defines our relation R. □

We will now prove that our relation does not admit a uniformisation in $\mathbf{FO}[<]$. For this, we count the number of candidates in R for each word of $\mathbb{A}^+$. Remember that a candidate for a word w in $\mathbb{A}^+$ is a word $\sigma \in \mathbb{B}^+$ such that $\left(\begin{smallmatrix}w\\\sigma\end{smallmatrix}\right) \in R$.

Let p, q be two positive natural numbers, and let w be the word $a^p \cdot b^q$. It is immediate to see that the number of candidates σ for w in R is: 0 if both p and q are odd; 1 if among p and q, one is even and the other is odd (and in this case, the only possible σ is $(0 \cdot 1)^{\frac{p}{2}} \cdot \lhd^q$ if for instance p is even); or 2 if both p and q are even (and in this case, the two possible candidates σ are $(0 \cdot 1)^{\frac{p}{2}} \cdot \lhd^q$ and $\rhd^p \cdot (0 \cdot 1)^{\frac{q}{2}}$).

The intuition is that $\mathrm{FO}[<]$ is too weak to express parity and therefore to distinguish these crucial cases, and this implies that a uniformisation of R cannot be definable in $\mathrm{FO}[<]$.

Claim 2.8. *R does not admit any* $\mathrm{FO}[<]$*-definable uniformisation.*

Proof. To prove this claim properly, let $F \subseteq \left(\begin{smallmatrix}\mathbb{A}\\\mathbb{B}\end{smallmatrix}\right)^+$ be a relation between finite words, definable in $\mathrm{FO}[\left(\begin{smallmatrix}\mathbb{A}\\\mathbb{B}\end{smallmatrix}\right), <]$. Let us assume, in order to get a contradiction, that F is a uniformisation of R.

Let S_F be the syntactic semigroup of F, and let $H \subseteq S_F$ be such that $F = h_{S_F}^{-1}(H)$, where $h_{S_F} \colon \left(\begin{smallmatrix}\mathbb{A}\\\mathbb{B}\end{smallmatrix}\right)^+ \to S_F$ is the syntactic homomorphism of F.

We define n as the natural number $\sharp(S_F) \geq 1$, and three words: $w = a^{2n} \cdot b^{2n}$ over the alphabet $\mathbb{A}$, and $\sigma_1 = (0 \cdot 1)^n \cdot \lhd^{2n}$, $\sigma_2 = \rhd^{2n} \cdot (0 \cdot 1)^n$ over the alphabet $\mathbb{B}$. From the remark which we wrote before the claim, we know that σ_1 and σ_2 are the only two candidates for w in R, and therefore, since F is a uniformisation of R, we have $\left(\begin{smallmatrix}w\\\sigma_1\end{smallmatrix}\right) \in F$ if and only if $\binom{w}{\sigma_2} \notin F$ (see Figure 2.3). We prove that, in fact, this equivalence does not hold, which will conclude the proof.

If the above equivalence holds, then it means that among the two pairs $\binom{w}{\sigma_1}$ and $\binom{w}{\sigma_2}$, one must not be in F. Without loss of generality, we suppose that $\binom{w}{\sigma_1} \notin F$. Then, we consider the two words $w' = w \cdot b = a^{2n} \cdot b^{2n+1}$ and $\sigma_1' = \sigma_1 \cdot \lhd = (0 \cdot 1)^n \cdot \lhd^{2n+1}$ (again, see Figure 2.3). Since w' is composed of an even number of a's before an odd number of b's, σ_1' is its only candidate in R.

$$\begin{array}{rl} w: & \overbrace{a\ \ a\ \cdots\ a\ \ a}^{2n}\ \overbrace{b\ \ b\ \cdots\ b\ \ b}^{2n} \\ \sigma_1: & 0\ \ 1\ \cdots\ 0\ \ 1\ \ \lhd\ \ \lhd\ \cdots\ \lhd\ \ \lhd \end{array} \quad \notin F$$

$$\begin{array}{rl} w: & \overbrace{a\ \ a\ \cdots\ a\ \ a}^{2n}\ \overbrace{b\ \ b\ \cdots\ b\ \ b}^{2n} \\ \sigma_2: & \rhd\ \ \rhd\ \cdots\ \rhd\ \ \rhd\ \ 0\ \ 1\ \cdots\ 0\ \ 1 \end{array} \quad \in F$$

$$\begin{array}{rl} w': & \overbrace{a\ \ a\ \cdots\ a\ \ a}^{2n}\ \overbrace{b\ \ b\ \cdots\ b\ \ b\ \ b}^{2n+1} \\ \sigma_1': & 0\ \ 1\ \cdots\ 0\ \ 1\ \ \lhd\ \ \lhd\ \cdots\ \lhd\ \ \lhd\ \ \lhd \end{array} \quad \notin F$$

Figure 2.3: If the first pair of words is not in F, then neither is the third pair.

Let $x = h_{S_F}(\binom{a \cdot a}{0 \cdot 1})$ and $y = h_{S_F}(\binom{b}{\lhd})$. Theorem 1.19 on page 44 tells us that S_F is aperiodic, and therefore we have $y^{2n} = y^{2n+1}$. Thus, we obtain:

$$h_{S_F}(\binom{w'}{\sigma_1'}) = x^n \cdot y^{2n+1} = x^n \cdot y^{2n} = h_{S_F}(\binom{w}{\sigma_1}) \notin H.$$

Hence, $\binom{w'}{\sigma_1'} \notin F$. Considering that σ_1' is the unique candidate for w' in R, F is not a uniformisation of R, and we get our wanted contradiction.

This concludes the proof of Claim 2.8: R does not admit any uniformisation that is definable in $\mathrm{FO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$. $\square$

From Fact 2.7 and Claim 2.8, we deduce Proposition 2.1:

Proposition 2.1. *Over finite words,* $\mathbf{FO}[<]$ *does not uniformise* $\mathbf{FO}^2[\mathsf{s}]$.

2.1.2 The class $\mathbf{FO}[\mathsf{s}]$ does not uniformise the class $\mathbf{FO}^2[\,]$

In this second subsection, we prove Proposition 2.2, which states that there is no possibility to uniformise $\mathbf{FO}^2[\,]$ in $\mathbf{FO}[\mathsf{s}]$. Here again, it is enough to show that a particular relation in the former class does not admit any uniformisation in the latter.

Let us define a relation $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$, with $\mathbb{A} = \{a, b\}$ and $\mathbb{B} = \{\Diamond, \odot\}$, that contains exactly the pairs of words $\binom{w}{\sigma}$ such that:

- a letter a at any position in w implies a letter $\Diamond$ at the same position in σ;
- there exists exactly one position labelled by $\odot$ in σ.

The first condition is definable by the formula $\neg\exists x.\ \binom{a}{\odot}(x)$, while the second by the formula $\exists x.\ \binom{b}{\odot}(x) \wedge \forall y.\ \binom{b}{\odot}(y) \implies x = y$. Therefore:

Fact 2.9. *R is definable in* $\mathrm{FO}^2[\,]$.

It is immediate to see that a word in $\mathbb{A}^+$ admits as many candidates in R as it has positions labelled by b. The key is that an $\mathrm{FO}[\mathsf{s}]$ formula cannot distinguish these positions if there are too many a's between them.

Claim 2.10. *R does not admit any uniformisation in* $\mathbf{FO}[\mathsf{s}]$.

Proof. Let us assume that R admits a uniformisation $F \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ definable in $\mathrm{FO}[\mathsf{s}]$, in order to get a contradiction. Let S_F be its syntactic semigroup, h_{S_F} its syntactic homomorphism, $H \subseteq S_F$ such that $F = h_{S_F}^{-1}(H)$, and let $n = \sharp(S_F)$.

Consider the three words $w = a^n{\cdot}b{\cdot}a^n{\cdot}b{\cdot}a^n$, $\sigma_1 = \Diamond^n{\cdot}\Diamond{\cdot}\Diamond^n{\cdot}\odot{\cdot}\Diamond^n$, and $\sigma_2 = \Diamond^n{\cdot}\odot{\cdot}\Diamond^n{\cdot}\Diamond{\cdot}\Diamond^n$ (see Figure 2.4). It is clear that σ_1 and σ_2 are the only two candidates for w in R.

$$\begin{array}{r|l|}
w: & \overbrace{a \;\cdots\; \cdots\; a}^{n}\;\; b\;\; \overbrace{a \;\cdots\; \cdots\; a}^{n}\;\; b\;\; \overbrace{a \;\cdots\; \cdots\; a}^{n} \\
\sigma_1: & \Diamond \;\cdots\; \cdots\; \Diamond\;\; \Diamond\;\; \Diamond \;\cdots\; \cdots\; \Diamond\;\; \odot\;\; \Diamond \;\cdots\; \cdots\; \Diamond
\end{array}$$

$$\begin{array}{r|l|}
w: & \overbrace{a \;\cdots\; \cdots\; a}^{n}\;\; b\;\; \overbrace{a \;\cdots\; \cdots\; a}^{n}\;\; b\;\; \overbrace{a \;\cdots\; \cdots\; a}^{n} \\
\sigma_2: & \Diamond \;\cdots\; \cdots\; \Diamond\;\; \odot\;\; \Diamond \;\cdots\; \cdots\; \Diamond\;\; \Diamond\;\; \Diamond \;\cdots\; \cdots\; \Diamond
\end{array}$$

Figure 2.4: The word w and its two candidates σ_1 and σ_2 in R.

Let $e = h_{S_F}(\binom{a}{\Diamond}^n)$ (it is idempotent), $s = h_{S_F}(\binom{b}{\Diamond})$, and $t = h_{S_F}(\binom{b}{\odot})$. We have the equalities $h_{S_F}(\binom{w}{\sigma_1}) = e{*}s{*}e{*}t{*}e$ and $h_{S_F}(\binom{w}{\sigma_2}) = e{*}t{*}e{*}s{*}e$, and from Corollary 1.21 on page 44, we can deduce that $h_{S_F}(\binom{w}{\sigma_1}) = h_{S_F}(\binom{w}{\sigma_2})$. This equality implies that $\binom{w}{\sigma_1} \in F$ if and only if $\binom{w}{\sigma_2} \in F$, and this enters into contradiction with the assumption that F is a uniformisation of R. $\square$

From Fact 2.9 and Claim 2.10, we deduce Proposition 2.2:

Proposition 2.2. *Over finite words,* $\mathbf{FO}[\mathsf{s}]$ *does not uniformise* $\mathbf{FO}^2[\,]$.

2.1.3 The class $\mathbf{FO}^2[\mathsf{s}]$ does not uniformise the class $\mathbf{FO}^2[<] \cap \mathbf{FO}^2[\mathsf{s}]$

Finally, we focus on Proposition 2.3, the last negative result of this chapter. Once again, a study of a single relation is enough to deduce the proposition.

Let $\mathbb{A}$ and $\mathbb{B}$ be respectively the alphabets $\{a, b, c, d\}$ and $\{\rhd, \odot, \lhd\}$. We consider the relation R as the following language, over the product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$:

$$\begin{pmatrix}\mathbb{A}\\ \rhd\end{pmatrix}^* \cdot \begin{pmatrix}a\\ \odot\end{pmatrix} \cdot \begin{pmatrix}\mathbb{A} \setminus \{b\}\\ \odot\end{pmatrix}^* \cdot \begin{pmatrix}c\\ \odot\end{pmatrix} \cdot \begin{pmatrix}\mathbb{A}\\ \lhd\end{pmatrix}^*.$$

no b's

w:	$\cdots$	$\cdots$	$\cdots$	a	$\cdots$	$\cdots$	$\cdots$	c	$\cdots$	$\cdots$	$\cdots$
σ:	$\rhd$	$\cdots$	$\rhd$	$\odot$	$\odot$	$\cdots$	$\odot$	$\odot$	$\lhd$	$\cdots$	$\lhd$

Figure 2.5: A generic pair of words in the relation R.

Expressing the existence of the three consecutive factors of a pair in R (*i.e.* the factor with the $\rhd$'s, the factor with the $\odot$'s, and the factor with the $\lhd$'s) can be done either with $<$ or with s, using only two variables:

Fact 2.11. *R is definable in both* $\mathrm{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, <]$ *and* $\mathrm{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, \mathsf{s}]$.

Proof. We define, for x being a first-order variable, $\rhd^2(x)$ as being the $\mathrm{FO}^0[\binom{\mathbb{A}}{\mathbb{B}}]$ formula $\binom{a}{\rhd}(x) \vee \binom{b}{\rhd}(x) \vee \binom{c}{\rhd}(x) \vee \binom{d}{\rhd}(x)$. We define $\odot^2(x)$ and $\lhd^2(x)$ similarly.

We define also $\varphi^{\mathsf{s}}_{\mathrm{first}}(x)$ as the $\mathrm{FO}^1[\binom{\mathbb{A}}{\mathbb{B}}, \mathsf{s}]$ formula $\forall y.\ \neg\mathsf{s}(y, x)$, expressing that the free variable x is evaluated into the first position of the word. Symmetrically, the formula $\varphi^{\mathsf{s}}_{\mathrm{last}}(x) := \forall y.\ \neg\mathsf{s}(x, y)$ expresses that x is evaluated into the last position of the word.

Then, we can define R by the $\mathrm{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, \mathsf{s}]$ formula:

$$
\begin{aligned}
\varphi^{\mathsf{s}} :=\ & \\
& \forall x.\ \neg \tbinom{b}{\odot}(x) \\
\wedge\ & \exists x.\ \varphi^{\mathsf{s}}_{\text{first}}(x) \wedge \big(\rhd^2(x) \vee \tbinom{a}{\odot}(x) \big) \\
\wedge\ & \exists x.\ \varphi^{\mathsf{s}}_{\text{last}}(x) \wedge \big(\tbinom{c}{\odot}(x) \vee \lhd^2(x) \big) \\
\wedge\ & \forall x, y.\ \mathsf{s}(x,y) \implies \quad \big[\rhd^2(x) \implies \big(\rhd^2(y) \vee \tbinom{a}{\odot}(y) \big) \big] \\
& \qquad\qquad\qquad\qquad \wedge \big[\big(\tbinom{a}{\odot}(x) \vee \tbinom{d}{\odot}(x) \big) \implies \odot^2(y) \big] \\
& \qquad\qquad\qquad\qquad \wedge \big[\tbinom{c}{\odot}(x) \implies \big(\odot^2(y) \vee \lhd^2(y) \big) \big] \\
& \qquad\qquad\qquad\qquad \wedge \big[\lhd^2(x) \implies \lhd^2(y) \big].
\end{aligned}
$$

The definition of R by an $\mathrm{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, <]$ formula is similar:

$$
\begin{aligned}
\varphi^{<} :=\ & \\
& \forall x.\ \neg \tbinom{b}{\odot}(x) \\
\wedge\ & \exists x.\ \tbinom{a}{\odot}(x) \wedge \forall y{<}x.\ \rhd^2(y) \\
\wedge\ & \exists x.\ \tbinom{c}{\odot}(x) \wedge \forall y{>}x.\ \lhd^2(y) \\
\wedge\ & \forall x, y.\ x < y \implies \big[\odot^2(x) \implies \big(\odot^2(y) \vee \lhd^2(y) \big) \big] \\
& \qquad\qquad\qquad\quad \wedge \big[\lhd^2(x) \implies \lhd^2(y) \big].
\end{aligned}
$$

$\square$

Here again, in order to prove that R is not uniformisable in $\mathrm{FO}^2[\mathsf{s}]$, we count the number of candidates each word of $\mathbb{A}^+$ has in R. By definition, it is exactly the number of pairs of positions $x < y$ of w such that $w(x) = a$, $w(y) = c$, and w has no b's between x and y. The idea is that two variables are not enough to distinguish two such pairs of positions $x_1 < y_1$ and $x_2 < y_2$, and, therefore, the formalism $\mathrm{FO}^2[\mathsf{s}]$ is not capable of distinguishing these candidates:

Claim 2.12. *R does not admit any uniformisation definable in* $\mathrm{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, \mathsf{s}]$.

Proof. In order to get a contradiction, let us assume that there exists a uniformisation F of R definable in $\mathrm{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, \mathsf{s}]$. Let S_F be the syntactic semigroup of F, h_{S_F} its syntactic homomorphism, and $H \subseteq S_F$ such that $F = h_{S_F}^{-1}(H)$.

We set $n = \sharp(S)$, and define u as the word $(d^n \cdot b \cdot d^n \cdot c \cdot a \cdot d^n)^n$ over $\mathbb{A}$ (see Figure 2.6). The word u does not belong to $\Pi_{\mathbb{A}^+}(R)$, since it admits an occurrence of b in-between its occurrences of a and c. However, $u \cdot c \cdot a \cdot u$ does belong to $\Pi_{\mathbb{A}^+}(R)$. We will use this fact to get to a contradiction.

$$u\colon \left(\boxed{\overbrace{d \ \cdots \ d}^{n \text{ times}}\ \ b\ \ \overbrace{d\ \cdots\ d}^{n \text{ times}}\ \ c\ \ a\ \ \overbrace{d\ \cdots\ d}^{n \text{ times}}}\right)^n$$

Figure 2.6: The word u.

Let $w = u \cdot c \cdot a \cdot u \cdot u \cdot c \cdot a \cdot u$. We notice that w admits exactly two candidates in R, and we denote them σ_1 and σ_2 (see Figure 2.7). Thus, F being a uniformisation of R, we have $\binom{w}{\sigma_1} \in F$ if and only if $\binom{w}{\sigma_2} \notin F$. Without loss of generality, we can suppose that $\binom{w}{\sigma_2} \notin F$.

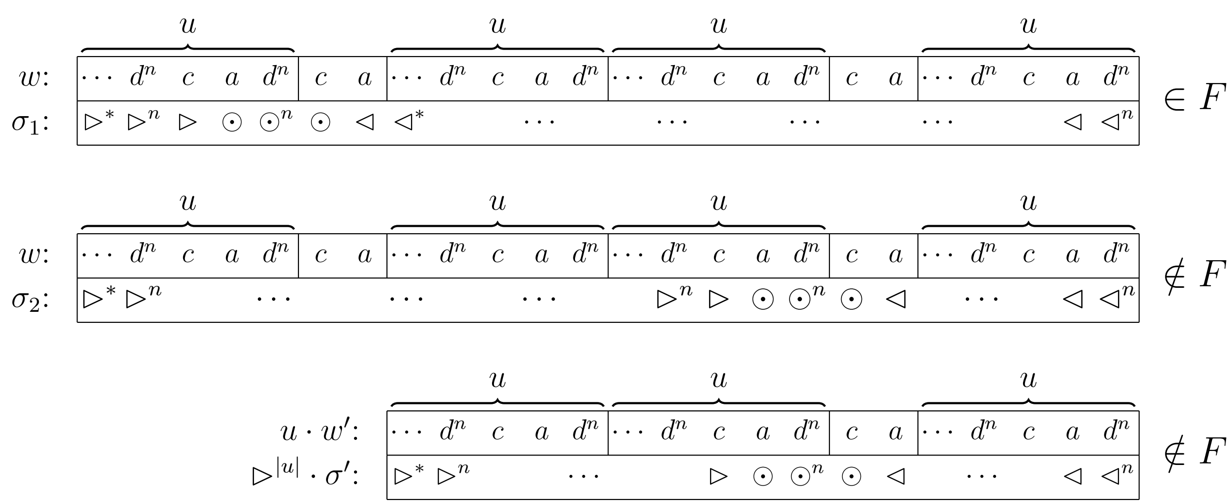


Figure 2.7: If the second pair is not in F, then neither is the third pair.

We consider now the word $w' = u \cdot c \cdot a \cdot u \in \mathbb{A}^+$. This time, this word admits a unique candidate σ' in R (again, see Figure 2.7). Moreover, $w = u \cdot c \cdot a \cdot u \cdot w'$, $\sigma_2 = \rhd^{|u|+2+|u|} \cdot \sigma'$, and $\rhd^{|u|} \cdot \sigma'$ is the unique candidate for $u \cdot w'$ in R.

Now, let $e = h_{S_F}(\binom{d}{\rhd}^n)$, let $x = h_{S_F}(\binom{b}{\rhd})$, and let $y = h_{S_F}(\binom{c \cdot a}{\rhd \cdot \rhd})$. We have $h_{S_F}(\binom{u}{\rhd^{|u|}}) = (e * x * e * y * e)^n$, and it follows:

$$
\begin{aligned}
h_{S_F}\left(\binom{u \cdot w'}{\rhd^{|u|} \cdot \sigma'}\right) &= h_{S_F}\left(\binom{u}{\rhd^{|u|}}\right) * h_{S_F}\left(\binom{w'}{\sigma'}\right) \\
&= (e * x * e * y * e)^n * h_{S_F}\left(\binom{w'}{\sigma'}\right) \\
&= (e * x * e * y * e)^n * y * (e * x * e * y * e)^n * h_{S_F}\left(\binom{w'}{\sigma'}\right) \\
&= h_{S_F}\left(\binom{u \cdot c \cdot a \cdot u}{\rhd^{|u|+2+|u|}}\right) * h_{S_F}\left(\binom{w'}{\sigma'}\right) \\
&= h_{S_F}\left(\binom{u \cdot c \cdot a \cdot u \cdot w'}{\rhd^{|u|+2+|u|} \cdot \sigma'}\right) \\
&= h_{S_F}\left(\binom{w}{\sigma_2}\right) \notin H,
\end{aligned}
$$

where the equality $(e * x * e * y * e)^n = (e * x * e * y * e)^n * y * (e * x * e * y * e)^n$ comes from Theorem 1.22 on page 44 (indeed, e is idempotent). This implies that $\binom{u \cdot w'}{\rhd^{|u|} \cdot \sigma'} \notin F$. This, together with the fact that $\rhd^{|u|} \cdot \sigma'$ is the unique candidate for $u \cdot w'$ in R, contradicts our hypothesis that F is a uniformisation of R. □

Once again, from Fact 2.11 and Claim 2.12, we deduce Proposition 2.3:

Proposition 2.3. *Over finite words,* $\mathbf{FO}^2[<, \mathsf{s}]$ *does not uniformise* $\mathbf{FO}^2[<] \cap \mathbf{FO}^2[\mathsf{s}]$.

2.2 The expressive power of FO[ ]

In this section, we study the expressive power of FO[ ], to give an overview of which relations can be defined in this formalism. We give two characterisations of this class. First, a semantic one: we show that FO[ ] is exactly capable of counting the letters up to some threshold. Second, an algebraic one, in terms of syntactic semigroups. These characterisations are not entirely new, but, since the formalism is mostly considered as somewhat weak, it is barely studied, and there is no reliable reference to them. It is worth citing Pin's lecture notes [Pin20], whose subsection 1.2 of Chapter XIV speaks about *commutative languages*, but yet without explicit mention of the formalism FO[ ] in itself. Thus, we give these characterisations for the sake of completeness.

2.2.1 Semantic characterisation of FO[ ]

In a first step, we formalise this intuition of "counting the letters up to some threshold" with notions of $\langle \mathbb{A}, N \rangle$-*maps*, and show that languages in **FO**[ ] can be described using such maps.

We consider, for each natural number N, a specific new symbol, $\trianglerighteq_N$. If $\mathbb{A}$ is an alphabet, and N is a natural number, then an $\langle \mathbb{A}, N \rangle$-*map* is a function from the alphabet $\mathbb{A}$ to the set $N \sqcup \{\trianglerighteq_N\} = \{0, 1, \ldots, N-1, \trianglerighteq_N\}$.

Naturally, since $\mathbb{A}$ and the set $N \sqcup \{\trianglerighteq_N\}$ are both finite, so is the set of $\langle \mathbb{A}, N \rangle$-maps:

Fact 2.13. *For each natural number N and each alphabet $\mathbb{A}$, there exists a finite number of $\langle \mathbb{A}, N \rangle$-maps.*

For each $\langle \mathbb{A}, N \rangle$-map m, we define the language L_m as the language of words in which the occurrences of each letter a match with the value of $m(a)$, the symbol $\trianglerighteq_N$ being interpreted as "at least N occurrences". More formally, L_m is the language

$$\left\{ \begin{array}{ll} w \in \mathbb{A}^+ \mid \text{for each } a \in N, & \text{if } m(a) \in N \text{ then } |w|_a = m(a) \\ & \text{if } m(a) = \trianglerighteq_N \text{ then } |w|_a \geq N \end{array} \right\}.$$

We recall that $|w|_a$ denotes the number of occurrences of the letter a in the word w (see the definition on page 25).

Example 2.14. *Let a, b, c be three distinct letters, and let m be the $\langle \mathbb{A}, 2 \rangle$-map that maps a to 1, b to $\trianglerighteq_2$, and c to 0. Then, its corresponding language is the set of words in $\{a, b, c\}^+$ that have exactly one occurrence of a, no occurrences of c, and at least two occurrences of b.*

If m is an $\langle \mathbb{A}, N \rangle$-map, then we denote by $\mathbb{A}_N(m)$ the sets of letters of $\mathbb{A}$ mapped to some natural number $i < N$ by m, and by $\mathbb{A}_{\trianglerighteq_N}(m)$ the sets of letters of $\mathbb{A}$ mapped to $\trianglerighteq_N$ by m.

First, we shall notice that two distinct $\langle \mathbb{A}, N \rangle$-maps induce two disjoint languages:

Remark 2.15. *If m_1 and m_2 are two distinct $\langle \mathbb{A}, N \rangle$-maps, then the languages L_{m_1} and L_{m_2} are disjoint.*

Proof. Let a be a letter such that $m_1(a) \neq m_2(a)$. It is immediate to see that for no words $w \in \mathbb{A}^+$, the value of $|w|_a$ can match with both $m_1(a)$ and $m_2(a)$: if $m_1(a)$ and $m_2(a)$ are both natural numbers smaller than N, then naturally $|w|_a$ cannot be equal to both $m_1(a)$ and $m_2(a)$, and if $m_1(a) = \trianglerighteq_N$ and $m_2(a) = n < N$ (or the contrary), then $|w|_a$ cannot be both equal to n and greater or equal to N. Therefore $L_{m_1} \cap L_{m_2} = \emptyset$. □

We define now $\mathbf{M}_N[\mathbb{A}]$ as the class of finite languages over $\mathbb{A}$ that can be written as a union of languages induced from $\langle \mathbb{A}, N\rangle$-maps: $L \in \mathbf{M}_N[\mathbb{A}]$ if $L = \bigsqcup_{m\in\Lambda} L_m$, with Λ being a set of $\langle \mathbb{A}, N\rangle$-maps. We may abuse the definition and say that L is obtained as a union of $\langle \mathbb{A}, N\rangle$-maps.

Our aim now is to prove that the class $\mathbf{FO}[\mathbb{A}]$ is exactly the union of all the $\mathbf{M}_N[\mathbb{A}]$'s, for $N \in \mathbb{N}$. First, let us notice that FO[ ] is capable of counting letters. We recall that $\mathrm{FO}_N[\mathbb{A}]$, which we denote here by $\mathrm{FO}_{d=N}[\mathbb{A}]$, is the class of $\mathrm{FO}[\mathbb{A}]$ formulae of quantifier depth at most N (see the definition on page 34 of the quantifier depth of a formula).

Fact 2.16. *Let a be a letter of an alphabet $\mathbb{A}$ and let $n \in \mathbb{N}$. There exist two formulae $\varphi_a^{=n}$ and $\varphi_a^{\geq n}$, respectively in $\mathrm{FO}_{d=n+1}[\mathbb{A}]$ and $\mathrm{FO}_{d=n}[\mathbb{A}]$ such that for all $w \in \mathbb{A}^+$:*
- *$w \vDash \varphi_a^{=n}$ if and only if $|w|_a = n$;*
- *$w \vDash \varphi_a^{\geq n}$ if and only if $|w|_a \geq n$.*

Proof. We define $\varphi_a^{\geq n}$ as

$$\exists x_0, \ldots, x_{n-1}.\ \bigwedge_{i\in j\in n} x_i \neq x_j \wedge \bigwedge_{i\in n} a(x_i)$$

and $\varphi_a^{=n}$ as the formula $\varphi^{\geq n} \wedge \neg\varphi^{\geq n+1}$. □

Using this lemma, we can prove the correlation between FO[ ] and maps. In the following proposition, $\mathrm{FO}^{k=N}[\mathbb{A}]$ is another notation for $\mathrm{FO}^N[\mathbb{A}]$:

Proposition 2.17. *For each alphabet $\mathbb{A}$ and each natural number N, the classes $\mathbf{M}_N[\mathbb{A}]$, $\mathbf{FO}_{d=N}[\mathbb{A}]$, and $\mathbf{FO}^{k=N}[\mathbb{A}]$ are the same.*

Proof. We begin by the inclusion $\mathbf{FO}_{d=N}[\mathbb{A}] \subseteq \mathbf{FO}^{k=N}[\mathbb{A}]$, which is immediate: in general, for any signature Σ, a formula φ in $\mathrm{FO}_{d=N}[\Sigma]$ is easily equivalent to a formula in $\mathrm{FO}^{k=N}[\Sigma]$. To show this, it suffices to rename by x_i each variable of φ that is introduced by the quantifiers of depth i. For instance, if $\varphi(x, y)$ and $\psi(x, z)$ are two formulae over Σ, both of quantifier depth 0, then the sentence $\exists x.\ \big(\forall y.\ \varphi(x, y) \wedge \exists z.\ \psi(x, z)\big)$, which is in $\mathrm{FO}_{d=2}[\Sigma]$, but not in $\mathrm{FO}^{k=2}[\Sigma]$, is equivalent to the sentence $\exists x_0.\ \big(\forall x_1.\ \varphi(x_0, x_1) \wedge \exists x_1.\ \psi(x_0, x_1)\big)$, which is in $\mathrm{FO}^{k=2}[\Sigma]$.

It remains to show the inclusions $\mathbf{M}_N[\mathbb{A}] \subseteq \mathbf{FO}_{d=N}[\mathbb{A}]$ and $\mathbf{FO}^{k=N}[\mathbb{A}] \subseteq \mathbf{M}_N[\mathbb{A}]$.

For the former inclusion, we rely on the formulae defined in Fact 2.16. For each $\langle \mathbb{A}, N\rangle$-map m, the language L_m is defined by the formula:

$$\varphi_m := \bigwedge_{a\in\mathbb{A}_N(m)} \varphi^{=m(a)} \wedge \bigwedge_{a\in\mathbb{A}_{\trianglerighteq N}(m)} \varphi^{\geq N},$$

which is in $\mathrm{FO}_{d=N}[\mathbb{A}]$. Hence, if $L = \bigsqcup_{m\in\Lambda} L_m$ is in $\mathbf{M}_N[\mathbb{A}]$, then it is defined by the formula $\bigvee_{m\in\Lambda} \varphi_m$, and it is in $\mathbf{FO}_{d=N}[\mathbb{A}]$.

Now, we prove the inclusion $\mathbf{FO}^{k=N}[\mathbb{A}] \subseteq \mathbf{M}_N[\mathbb{A}]$.

We consider, for any $w \in \mathbb{A}^+$ and for any integer N, the following $\langle \mathbb{A}, N\rangle$-map:

$$m_N(w) = \begin{pmatrix} \mathbb{A} & \to & N \sqcup \{\trianglerighteq_N\} & \\ a & \mapsto & |w|_a & \text{if } |w|_a < N \\ & & \trianglerighteq_N & \text{if } |w|_a \geq N \end{pmatrix}.$$

Notice that, if w_1 and w_2 are two finite words over $\mathbb{A}$, then the functions $m_N(w_1)$ and $m_N(w_2)$ coincide (meaning that they map the same letters to the same elements) if and only if for all $a \in \mathbb{A}$, if $|w_1|_a \neq |w_2|_a$ then $\min\{|w_1|_a, |w_2|_a\} \geq N$. We denote this condition by $w_1 \sim_N w_2$.

For all $w_1, w_2 \in \mathbb{A}^+$, we consider, for all $d \in \mathbb{N}$, the game $\mathcal{G}_d^{\mathrm{FO}^N[\mathbb{A}]}(w_1, w_2)$ introduced on page 36, in Section 1.2. This is the Ehrenfeucht-Fraïssé game for First-Order Logic that is played during d turns over w_1 and w_2, and with $2N$ tokens, each number $j \in N$ being on exactly two of them. After each turn, some tokens are placed on the positions of each word, and Duplicator wins the game if at each turn, the related tokens (meaning with the same number on them) are on positions of different words, labelled by the same letter of $\mathbb{A}$.

It immediately appears that $w_1 \sim_N w_2$ if and only if $w_1 \equiv^N w_2$, meaning if and only if for every $d \in \mathbb{N}$, Duplicator has a winning strategy for the game $\mathcal{G}_d^{\mathrm{FO}^N[\mathbb{A}]}(w_1, w_2)$.

Indeed, let us suppose that $w_1 \sim_N w_2$, and let us say that on some turn of the game, Spoiler places a token on a position x of w_1, on which there was no token yet, and let a be the letter $w_1(x)$. If $|w_1|_a \geq N$, then also $|w_2|_a \geq N$, and Duplicator can move the corresponding token to a new position y of w_2 also labelled by a. On the contrary, if $|w_1|_a < N$, then $|w_1|_a = |w_2|_a$, and, the same way, Duplicator can select a new position of w_2 labelled by a. With this strategy, Duplicator always ensures that corresponding tokens are placed on positions identically labelled, and eventually wins the game.

Reciprocally, let us suppose that $w_1 \nsim_N w_2$. It means that there is some letter a in $\mathbb{A}$ such that $\min\{|w_1|_a, |w_2|_a\} < N$ and $|w_1|_a \neq |w_2|_a$. Without loss of generality, let us suppose that $\min\{|w_1|_a, |w_2|_a\} = |w_1|_a$. The strategy for Spoiler is to put on every turn a new token on a previously unselected position of w_2 labelled by a. On turn $|w_1|_a$, Duplicator has no more labelled-by-a positions to select in $\mathrm{Dom}(w_1)$, and eventually loses the game $\mathcal{G}_d^{\mathrm{FO}^N[\mathbb{A}]}(w_1, w_2)$, with the number of turns being $d = |w_1|_a+1$.

Let L be a language of finite words over $\mathbb{A}$, defined by an $\mathrm{FO}^{k=N}[\mathbb{A}]$ formula. From

Theorem 1.9 and the observation above, we know now that for all $w_1 \sim_N w_2$, we have $w_1 \in L$ if and only if $w_2 \in L$. Therefore, L is exactly the union of all the L_m's, $m \in \Lambda$, where Λ is the set $\{m_N(w) \mid w \in L\}$, and we have proven the final inclusion. $\square$

An immediate consequence of this characterisation is that the classes $\mathbf{M}_N[\mathbb{A}]$ inherit the closures of First-Order Logic:

Corollary 2.18. *For any alphabet $\mathbb{A}$ and natural numbers N, N':*
- $\mathbf{M}_N[\mathbb{A}]$ *is closed under Boolean combinations,*
- *if $N \leq N'$ then $\mathbf{M}_N[\mathbb{A}] \subseteq \mathbf{M}_{N'}[\mathbb{A}]$.*

Notice that these two properties were directly deducible from definitions of maps and of the languages L_f, but a formal proof of it would have been less straightforward.

Another consequence of Proposition 2.17 is that belonging to the classes $\mathbf{M}_N[\mathbb{A}]$ is decidable:

Corollary 2.19. *There exists an algorithm that inputs an* MSO$[\mathbb{A}, <]$ *sentence, with $\mathbb{A}$ being any alphabet and N any natural number, and outputs a set Λ of $\langle \mathbb{A}, N\rangle$-maps such that $\mathcal{L}(\varphi) = \bigsqcup_{m\in\Lambda} L_m$ if there exists such, and* **NO** *if $\mathcal{L}(\varphi)$ does not belong to $\mathbf{M}_N[\mathbb{A}]$.*

Proof. Recall from Corollary 1.17 on Page 43 that there exists an algorithm that tells whether one regular language is included in another. We can deduce from it an algorithm that inputs two formulae and tests whether their corresponding languages are equal.

Fact 2.13 tells us that there exists a finite number of $\langle \mathbb{A}, N\rangle$-maps, and therefore there exists a finite number of sets of $\langle \mathbb{A}, N\rangle$-maps. Since we know now from Proposition 2.17 an MSO$[\mathbb{A}]$ formula ψ_Λ that defines the language $\bigsqcup_{m\in\Lambda} L_m$ for each of these subsets Λ, we simply have to test the equality between $\mathcal{L}(\varphi)$ and all of these languages $\mathcal{L}(\psi_\Lambda)$. $\square$

There may exist more efficient methods. However, since we focus on decidability issues in this chapter, we content ourselves with this naive algorithm.

Before concluding this subsection, we show some basic results about projections of languages written as unions of $\langle \binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-maps, with $\mathbb{A}$ and $\mathbb{B}$ being two alphabets. In such cases, the maps are more likely denoted by the letter r, and the languages L_r are rather written R_r, to emphasise that we see them as relations.

Let r be an $\langle \binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-map, and let $a \in \mathbb{A}$. We define a way of partitioning the alphabet $\mathbb{B}$. We define $\mathbb{B}^a_N(r)$ as the set $\{b \in \mathbb{B} \mid r(\binom{a}{b}) \in N\}$, and $\mathbb{B}^a_{\trianglerighteq_N}(r)$ as the set $\{b \in \mathbb{B} \mid r(\binom{a}{b}) = \trianglerighteq_N\}$, in such a way that $\mathbb{B} = \mathbb{B}^a_N(r) \sqcup \mathbb{B}^a_{\trianglerighteq_N}(r)$. Finally, we define $r^*(a)$ as being the natural number $\Sigma_{b\in\mathbb{B}^a_N(r)} r(\binom{a}{b})$.

Lemma 2.20. *Let $N \in \mathbb{N}$ and let r be an $\langle \binom{\mathbb{A}}{\mathbb{B}}, N \rangle$-map, with $\mathbb{A}$ and $\mathbb{B}$ being two alphabets. Then, a word w in $\mathbb{A}^+$ belongs to $\Pi_{\mathbb{A}^+}(R_r)$ if and only if for all letters $a \in \mathbb{A}$:*
- *if $\mathbb{B}^a_{\trianglerighteq N}(r) = \emptyset$ then $|w|_a = r^*(a)$,*
- *if $\mathbb{B}^a_{\trianglerighteq N}(r) \neq \emptyset$ then $|w|_a \geq r^*(a) + N \times |\mathbb{B}^a_{\trianglerighteq N}(r)|$.*

Proof. This lemma is simply about counting.

Let $w \in \mathbb{A}^+$ and let us suppose that there exists some $\sigma \in \mathbb{B}^+$ such that $\binom{w}{\sigma} \in R_r$. Let $a \in \mathbb{A}$. We have the equality:

$$\begin{array}{rcll} |w|_a & = & \Sigma_{b \in \mathbb{B}} |\binom{w}{\sigma}|_{\binom{a}{b}} & \\ & = & \Sigma_{b \in \mathbb{B}^a_N(r)} |\binom{w}{\sigma}|_{\binom{a}{b}} & + \Sigma_{b \in \mathbb{B}^a_{\trianglerighteq N}(r)} |\binom{w}{\sigma}|_{\binom{a}{b}} \\ & = & r^*(a) & + \Sigma_{b \in \mathbb{B}^a_{\trianglerighteq N}(r)} |\binom{w}{\sigma}|_{\binom{a}{b}}. \end{array}$$

Hence, if $\mathbb{B}^a_{\trianglerighteq N}(r) = \emptyset$ then $|w|_a = r^*(a)$, and if $\mathbb{B}^a_{\trianglerighteq N}(r) \neq \emptyset$ then $|w|_a \geq r^*(a) + N \times |\mathbb{B}^a_{\trianglerighteq N}(r)|$.

Reciprocally, if $w \in \mathbb{A}^+$ is such that the two implications hold for each $a \in \mathbb{A}$, then we can easily construct some word $\sigma \in \mathbb{B}^+$ such that $\binom{w}{\sigma} \in R_r$, in the following way. For each $a \in \mathbb{A}$, let us rename $\mathbb{B}^a_N(r)$ as $\{b_0, \ldots, b_{k-1}\}$, and $\mathbb{B}^a_{\trianglerighteq N}(r)$ as $\{b_k, \ldots, b_{m-1}\}$. Then σ maps to b_0 the first $r(\binom{a}{b_0})$ labelled-by-a positions of w, to b_1 the next $r(\binom{a}{b_1})$ such positions, and so on, up to b_{k-1}. Then, σ maps to b_k the eventual next N labelled-by-a positions of w, and so on up to b_{m-2}. Finally, the remaining labelled-by-a positions of w are mapped to b_{m-1} by σ.

Our construction ensures that for each $\binom{a}{b} \in \binom{\mathbb{A}}{\mathbb{B}}$, if $r(\binom{a}{b}) \in N$ then $|\binom{w}{\sigma}|_{\binom{a}{b}}$ is $r(\binom{a}{b})$, and if $r(\binom{a}{b})$ is $\trianglerighteq_N$ then $|\binom{w}{\sigma}|_{\binom{a}{b}} \geq N$, and we can conclude. □

The direct corollary of this lemma is that the projection of a map is also a map:

Corollary 2.21. *Let R be a binary relation of finite words, over a product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$. We suppose that $R \in \mathbf{M}_N[\binom{\mathbb{A}}{\mathbb{B}}]$, for some $N \in \mathbb{N}$. Then, $\Pi_{\mathbb{A}^+}(R)$, the projection of R onto $\mathbb{A}^+$, is in $\mathbf{M}_{N \times |\mathbb{B}|}[\mathbb{A}]$.*

Proof. It is a direct consequence of Lemma 2.20, and of the fact that for every letter a in $\mathbb{A}$, the natural number $r^*(a)$ is smaller than $N \times |\mathbb{B}^a_{\trianglerighteq N}(r)|$ and $r^*(a) + N \times |\mathbb{B}^a_{\trianglerighteq N}(r)|$ is at most $N \times |\mathbb{B}|$. □

2.2.2 Algebraic characterisation of FO[ ]

After having described the expressive power of the class **FO**[ ] in terms of maps in the previous subsection, we characterise here its languages as those whose syntactic semigroups satisfy some constraints. These constraints being finiteness, aperiodicity, and *commutativity.*

A semigroup such that $s * t = t * s$ for all $s, t \in S$ is called *commutative*. We recall that a finite semigroup is called aperiodic if for all $s \in \mathbb{A}$, we have $s^\sharp * s = s^\sharp$ (see the definition on page 44).

Theorem 2.22. *Let $L \subseteq \mathbb{A}^+$ be a language of finite words, over an alphabet $\mathbb{A}$. Then L is definable in* FO[$\mathbb{A}$] *if and only if its syntactic semigroup S_L if finite, aperiodic, and commutative. Moreover in the latter case, L is definable by a formula in* FO[$\mathbb{A}$] *of depth at most $\sharp(S_L)$.*

Proof. To prove the "only if" implication, we recall the definition of the syntactic semigroup of L, defined on page 42. It is obtained from the equivalence relation $\equiv_L$ over $\mathbb{A}^+$ defined by $u \equiv_L v$ if for all $w_1, w_2 \in \mathbb{A}^*$, $w_1 \cdot u \cdot w_2 \in L$ if and only if $w_1 \cdot v \cdot w_2 \in L$. To each $u \in \mathbb{A}^+$ corresponds its equivalence class $[u]_{\equiv_L} = \{v \in \mathbb{A}^+ \mid u \equiv_L v\}$, and S_L is the set $\{[u]_{\equiv_L} \mid u \in \mathbb{A}^+\}$ of these equivalence classes, with the product $[u]_{\equiv_L} * [v]_{\equiv_L} = [u \cdot v]_{\equiv_L}$, and the syntactic homomorphism h_{S_L} is defined by $h_{S_L}(u) = [u]_{\equiv_L}$.

Let us suppose that L is in **FO**[$\mathbb{A}$]. By Proposition 2.17, L can be written as $\bigsqcup_{m \in \Lambda} L_m$, where Λ is a set of $\langle \mathbb{A}, N \rangle$-map, N being the quantifier depth of a formula defining L. By Theorem 1.19, we know that S_L is finite and aperiodic. It remains to prove that it is commutative.

This is rather straightforward: let $s, t \in S_L$. These elements can be written as $s = [u]_{\equiv_L}$ and $t = [v]_{\equiv_L}$, with u and v being two finite words over $\mathbb{A}$, and we have $s * t = [u \cdot v]_{\equiv_L}$, $t * s = [v \cdot u]_{\equiv_L}$. Therefore, we have to prove that $u \cdot v \equiv_L v \cdot u$, which is immediate: for all $w_1, w_2 \in \mathbb{A}^*$, $w_1 \cdot u \cdot v \cdot w_2$ and $w_1 \cdot v \cdot u \cdot w_2$ contain exactly the same letters, and therefore, for all $m \in \Lambda$, $w_1 \cdot u \cdot v \cdot w_2 \in L_m$ if and only if $w_1 \cdot v \cdot u \cdot w_2 \in L_m$, which means that $w_1 \cdot u \cdot v \cdot w_2$ and $w_1 \cdot v \cdot u \cdot w_2$ equivalently belong to L, and therefore $u \cdot v \equiv_L v \cdot u$: $s * t = t * s$, and S_L is commutative. We have proven one implication.

Now, we prove the "if" implication. Let us suppose that L is recognised by some finite, aperiodic, and commutative semigroup S, via a homomorphism h.

We suppose without loss of generality that $\mathbb{A}$ is the set $\{a_0, \ldots, a_{k-1}\}$. For any finite word $w \in \mathbb{A}^+$, we have $h(w) = h\big(w(0)\big) * h\big(w(1)\big) * \cdots * h\big(w(|w|-1)\big)$, and, because S is commutative, we have $h(w) = h(a_0)^{|w|_{a_0}} * h(a_1)^{|w|_{a_1}} * \cdots * h(a_{k-1})^{|w|_{a_{k-1}}}$ (if $|w|_{a_i} = 0$ for some i, we can simply ignore it by defining $s * h(a_i)^0 = h(a_i)^0 * s = s$, and this will not be a problem since there are necessarily other i's such that $|w|_{a_i} \neq 0$).

We define $N = \sharp(S)$. By the aperiodicity of S, we have, for each $s \in S$, and all $i \in k$ such that $|w|_{a_i} \geq N$, $h(a_i)^{|w|_{a_i}} = h(a_i)^N$. Therefore, for each $\langle \mathbb{A}, N\rangle$-map m, if w_1 and w_2 are in L_m, then $h(w_1) = h(w_2)$, because $h(a_i)^{|w_1|_{a_i}} = h(a_i)^{|w_2|_{a_i}}$ for all $i \in k$. Hence, $w_1 \in L$ if and only if $w_2 \in L$, and we have $L = \bigsqcup_{w \in L} L_{m_N(w)}$, where $m_N(w)$ is the $\langle \mathbb{A}, N\rangle$-map defined as in the proof of Proposition 2.17: it maps each $a \in \mathbb{A}$ to $|w|_a$ if $|w|_a < N$, and to $\trianglerighteq_N$ if $|w|_a \geq N$. $\square$

We can deduce from this algebraic characterisation that belonging to the class **FO**[ ] is decidable:

Corollary 2.23. *There exists an algorithm that inputs an* MSO$[\mathbb{A}, <]$ *sentence, with* $\mathbb{A}$ *being any alphabet, and outputs a natural number* N *and a set* Λ *of* $\langle \mathbb{A}, N\rangle$*-maps such that* $\mathcal{L}(\varphi)$ *is* $\bigsqcup_{m \in \Lambda} L_m$ *if there exist some, and* **NO** *if* $\mathcal{L}(\varphi)$ *does not belong to* **FO**$[\mathbb{A}]$.

Proof. From Proposition 1.15 on page 42, we can compute the syntactic semigroup S_L of $L = \mathcal{L}(\varphi)$. Since it is finite, we can compute $\sharp(S_L)$ and test if the semigroup is aperiodic and commutative. If it is not, we output **NO**. If it is, we input φ and $N := \sharp(S_L)$ in the algorithm of Corollary 2.19, to obtain the $\langle \mathbb{A}, N\rangle$-maps defining L. $\square$

2.3 Uniformising FO[ ] in FO[<]

Now, we use the characterisation of **FO**[ ], proven in the previous subsection, to show how to uniformise in FO[<] relations definable in FO[ ].

First, we prove that **FO**[<] can uniformise each individual relation R_r, for any $\big\langle \big(\begin{smallmatrix}\mathbb{A}\\ \mathbb{B}\end{smallmatrix}\big), N\big\rangle$-map r:

Lemma 2.24. *Let* $N \in \mathbb{N}$ *and let* r *be an* $\big\langle \big(\begin{smallmatrix}\mathbb{A}\\ \mathbb{B}\end{smallmatrix}\big), N\big\rangle$*-map, with* $\mathbb{A}$ *and* $\mathbb{B}$ *being two alphabets. Then* R_r *admits a uniformisation in* **FO**[<].

Proof. Without loss of generality, we can assume that $\mathbb{B}$ is the alphabet $\{b_0, b_1, \ldots, b_{m-1}\}$, with $m = |\mathbb{B}|$. To define a particular candidate σ in R_r for each word $w \in \Pi_{\mathbb{A}^+}(R_r)$, we repeat the strategy we used in the proof of Lemma 2.20: we divide, for each letter $a \in \mathbb{A}$, the set $\{x \in \mathrm{Dom}(w) \mid w(x) = a\}$ into disjoint subsets $X_0 < X_1 < \ldots < X_{m-1}$, and to ensure that each position of σ is labelled by b_j when it is in X_j.

Let us consider a letter a of $\mathbb{A}$. Again, without loss of generality, we can assume that the b_j's are such that the first k are in $\mathbb{B}^a_N(r)$, and that the last $m-k$ are in $\mathbb{B}^a_{\trianglerighteq N}(r)$, for some k. Then, we define a finite word u_a over $\mathbb{B}$ as $b_0^{r(\binom{a}{b_0})} \cdots b_{k-1}^{r(\binom{a}{b_{k-1}})} \cdot b_k^N \cdots b_{m-1}^N$, like depicted on Figure 2.8:

$$u_a\colon\ \boxed{\overbrace{b_0 \quad \cdots \quad b_0}^{r(\binom{a}{b_0})\text{ times}} \cdots \overbrace{b_{k-1} \quad \cdots \quad b_{k-1}}^{r(\binom{a}{b_{k-1}})\text{ times}}\ \overbrace{b_k \quad \cdots \quad b_k}^{N\text{ times}} \cdots \overbrace{b_{m-1} \quad \cdots \quad b_{m-1}}^{N\text{ times}}}$$

Figure 2.8: The word u_a used in the proof of Lemma 2.24.

The idea is to associate to the i-th a in w the i-th letter in u_a, with the following FO[<] formula (n being the length of u_a):

$$\begin{aligned}
&\varphi_{r,a} := \\
&\quad \exists x_0, \ldots, x_{n-1}. \bigwedge_{j \in n} \binom{a}{u_a(j)}(x_j) \wedge \bigwedge_{j \in n-1} x_j < x_{j+1} \\
&\qquad\qquad\qquad \wedge \forall x. \bigvee_{b \in \mathbb{B}} \binom{a}{b}(x) \\
&\qquad\qquad\qquad\qquad \implies \Big(\bigvee_{i \in n} x = x_i \vee \big(x_{n-1} < x \wedge \binom{a}{b_{m-1}}(x) \big) \Big).
\end{aligned}$$

In the case when $\mathbb{B}^a_{\trianglerighteq N}(r) = \emptyset$ (*i.e.* $k = m$), we replace the subformula $\Big(\bigvee_{i \in n} x = x_i \vee \big(x_{n-1} < x \wedge \binom{a}{b_{m-1}}(x) \big) \Big)$ by the simpler $\bigvee_{i \in n} x = x_i$.

If $\sigma \in \mathbb{B}^+$ is such that $\binom{w}{\sigma} \models \varphi_{r,a}$, then by construction we know that σ has, for each $b_i \in \mathbb{B}^a_N(r)$, exactly $r(\binom{a}{b_i})$ positions labelled by b_i, and for each $b_i \in \mathbb{B}^a_{\trianglerighteq N}$, at least N positions labelled by b_i. Moreover, any two such words σ and σ' agree on the positions labelled by a in w.

Now, we define φ_r as being the formula $\bigwedge_{a \in \mathbb{A}} \varphi_{r,a}$. From the paragraph above, we know that $\mathcal{L}(\varphi_r)$ is a uniformisation of R_r. $\square$

Now, we can prove Proposition 2.5 and conclude the section:

Proposition 2.5. *Over finite words, the class* **FO**[<] *uniformises the class* **FO**[ ].

Proof. Let $\mathbb{A}$ and $\mathbb{B}$ be two alphabets. Let R be a relation definable in FO[ ], over the product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$. Proposition 2.17 tells us that there exists a natural number N and a set Υ of $\langle\binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-maps such that $R = \bigsqcup_{r\in\Upsilon} R_r$.

Lemma 2.24 tells us that for each $r \in \Upsilon$, R_r admits a uniformisation $F_r = \mathcal{L}(\varphi_r)$, where φ_r is an FO[<] formula. The seemingly natural formula $\varphi := \bigvee_{r\in\Upsilon} \varphi_r$ does *a priori* not define a uniformisation of R, because some w's in $\Pi_{\mathbb{A}^+}(R)$ may appear in different $\Pi_{\mathbb{A}^+}(R_r)$'s.

However, Fact 2.21 tells us that for each $r \in \Upsilon$, there exists an FO[$\mathbb{A}$] sentence φ^r_{proj} over $\mathbb{A}$ defining $\Pi_{\mathbb{A}^+}(R_r)$. If we replace, in one formula φ^r_{proj}, all subformulae of the shape $a(x)$ by the corresponding formula $\bigvee_{b\in\mathbb{B}} \binom{a}{b}(x)$, we obtain a formula ${\varphi^r_{\text{proj}}}'$ such that $\binom{w}{\sigma} \models {\varphi^r_{\text{proj}}}'$ if and only if $w \in \Pi_{\mathbb{A}^+}(R_r)$.

Now, if we name the $\langle\binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-maps of Υ as $r_0, r_1, \ldots, r_{\ell-1}$, it is now possible, in FO[$\binom{\mathbb{A}}{\mathbb{B}}$], to map each $w \in \Pi_{\mathbb{A}^+}(R)$ to one of its candidates in R: if $w \in \Pi_{\mathbb{A}^+}(R_{r_0})$, then we choose the corresponding σ_0 associated by the uniformisation $\mathcal{L}(\varphi_{r_0})$; if it is not the case but if $w \in \Pi_{\mathbb{A}^+}(R_{r_1})$, we choose the corresponding σ_1 associated by the uniformisation $\mathcal{L}(\varphi_{r_1})$, and so on...

The corresponding FO[<] formula is ψ_0, where: $\psi_{\ell-1}$ is $\varphi_{r_{\ell-1}}$, and for $0 \leq i < \ell-1$, ψ_i is $\left({\varphi^{r_i}_{\text{proj}}}' \Rightarrow \varphi_{r_i}\right) \wedge \left(\neg{\varphi^{r_i}_{\text{proj}}}' \Rightarrow \psi_{i+1}\right)$.

We have shown that R admits a uniformisation in **FO**[<], which concludes the proof of the proposition. □

2.4 The decidability to uniformise a regular relation in FO[ ]

In this final section, we prove the decidability of uniformising regular relations in **FO**[ ]. The key is to prove that if a regular relation $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ admits a uniformisation in **FO**[$\binom{\mathbb{A}}{\mathbb{B}}$], then it must admit one in $\mathbf{M}_Z$[$\binom{\mathbb{A}}{\mathbb{B}}$], with Z being a natural number which we can compute from R.

We recall our remark on page 56 and emphasise that, in order to admit a uniformisation in **FO**[ ], the relation R needs not be definable in FO[ ] itself. For instance, if $\mathbb{A} = \{a\}$, and $\mathbb{B} = \{\odot, \Diamond\}$, then the relation $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ composed of the pair of words $\binom{w}{\sigma}$ such that σ contains an even number of occurrences of $\Diamond$ is not even definable in FO[$\binom{\mathbb{A}}{\mathbb{B}}$, <] (the fact that parity cannot be defined in FO[<] is proven similarly as Claim 2.8, via the algebraic characterisation of this formalism). Yet it admits as uniformisation the relation $F = \left\{\binom{w}{\sigma} \in \binom{\mathbb{A}}{\mathbb{B}}^+ \mid |\sigma|_\Diamond = 0\right\}$, which is in **FO**[$\binom{\mathbb{A}}{\mathbb{B}}$].

In Section 2.2, we decided to denote $\langle\binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-maps by the letter r, and their corresponding languages by R_r, to emphasise that we see them as relations. In this section, when we have the additional assumption that said relations are functional, we denote the maps by f and their corresponding languages by F_f. Finally, in this section, we denote $\langle \mathbb{A}, N\rangle$-maps by the letter m, their languages L_m's are domains of functions F_f's.

2.4.1 Basic properties of relational maps

We begin this section by enumerating some basic properties about relations and uniformisations in the class $\mathbf{M}_N[\binom{\mathbb{A}}{\mathbb{B}}]$.

Corollary 2.21 tells us that the projection of a language in $\mathbf{M}_N[\binom{\mathbb{A}}{\mathbb{B}}]$ is in $\mathbf{M}_{N\times|\mathbb{B}|}[\mathbb{A}]$. We obtain the same conclusion if a relation R is not in the class $\mathbf{M}_N[\binom{\mathbb{A}}{\mathbb{B}}]$ itself, but if it admits some uniformisation in it:

Corollary 2.25. *Let R be a binary relation of finite words, over the product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$. We suppose that R admits a uniformisation $F \in \mathbf{M}_N[\binom{\mathbb{A}}{\mathbb{B}}]$, for some $N \in \mathbb{N}$. Then, $\Pi_{\mathbb{A}^+}(R)$, the projection of R onto $\mathbb{A}^+$, is in $\mathbf{M}_{N\times|\mathbb{B}|}[\mathbb{A}]$.*

Proof. By the definition of a uniformisation, we have $\Pi_{\mathbb{A}^+}(R) = \Pi_{\mathbb{A}^+}(F)$, which we know is in $\mathbf{M}_{N\times|\mathbb{B}|}[\mathbb{A}]$, by Corollary 2.21. □

We noticed in Remark 2.15 that distinct maps induce disjoint languages. Under the additional assumption that the union of these maps is functional, then we have a stronger conclusion, which is that also the projections are disjoint, as stated in the following lemma:

Lemma 2.26. *Let F be a binary relation of finite words, over a product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$. We suppose that F is functional, and also that it is in $\mathbf{M}_N[\binom{\mathbb{A}}{\mathbb{B}}]$: $F = \bigsqcup_{f\in\Upsilon} F_f$, with Υ being a set of $\langle\binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-maps. Then the $\Pi_{\mathbb{A}^+}(F_f)$'s, the projections of these F_f's onto $\mathbb{A}^+$, are pairwise disjoint.*

Proof. Suppose that there exists some $w \in \mathbb{A}^+$ being both in $\Pi_{\mathbb{A}^+}(F_{f_1})$ and $\Pi_{\mathbb{A}^+}(F_{f_2})$, with f_1 and f_2 being two distinct $\langle\binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-maps of Υ. By the definition of the projection, there exist two finite words σ_1 and σ_2, both over $\mathbb{B}$, such that $\binom{w}{\sigma_1} \in F_{f_1}$ and $\binom{w}{\sigma_2} \in F_{f_2}$. Since F_{f_1} and F_{f_2} are disjoint (by Remark 2.15), σ_1 and σ_2 must be distinct, and therefore w has two different candidates in F, which contradicts the assumption that F is functional. □

Notice that this lemma is not specific to $\langle\binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-maps: more generally, if F is a functional relation that can be written as $\bigsqcup_{i\in I} F_i$, with the F_i's being pairwise disjoint functional relations, then their projections must be disjoint too, by the same argument.

Finally, we show that maps that induce functional relations have some strong restrictions:

Lemma 2.27. *Let f be an $\langle\binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-map. Then the relation F_f is functional if and only if for all $a \in \mathbb{A}$, there exists at most one $b \in \mathbb{B}$ such that $f(\binom{a}{b}) \neq 0$.*

Proof. It is clear that a map f that satisfies this property induces a functional relation F_f: in this case, the letters in a word $w \in \Pi_{\mathbb{A}^+}(F_f)$ uniquely determine the letters of a word $\sigma \in \mathbb{B}^+$ such that $\binom{w}{\sigma} \in F_f$.

Now, let us suppose that F_f is functional. Let $w \in \Pi_{\mathbb{A}^+}(F_f)$ and let $\sigma \in \mathbb{B}^{|w|}$ such that $\binom{w}{\sigma} \in F_f$. If there exist two distinct letters b_1 and b_2 such that $f(\binom{a}{b_1}) \neq 0$ and $f(\binom{a}{b_2}) \neq 0$, then there exist two distinct positions x and y in $\mathrm{Dom}(w)$ such that $\binom{w}{\sigma}(x) = \binom{a}{b_1}$ and $\binom{w}{\sigma}(y) = \binom{a}{b_2}$. If we define σ' as the word over $\mathbb{B}$ having the same domain as σ, but that swaps the labels of x and y (meaning it maps x to b_2, y to b_1, and any other position $z \in \mathrm{Dom}(w)$ to $\sigma(z)$), then it is immediate to see that it still satisfies $\binom{w}{\sigma'} \in F_f$. Since $\sigma \neq \sigma'$, it contradicts the assumption that F_f is functional. Therefore, there cannot exist more than one such letter b. □

2.4.2 Inductive lemma and decidability theorem

In this second subsection, we prove the following crucial lemma, which we mentioned at the beginning of our section: if $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ admits some uniformisation in $\mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$, and if its projection also is in $\mathbf{FO}[\mathbb{A}]$, then we can compute some natural number Z such that R also admits a uniformisation in $\mathbf{M}_Z[\binom{\mathbb{A}}{\mathbb{B}}]$.

Lemma 2.28. *Let $\mathbb{A}$, $\mathbb{B}$ be two alphabets and let $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ be a regular relation. Then, for every natural number N, and every $\langle\mathbb{A}, N\rangle$-map m, if the following two assumptions are true:*

- *$L_m \subseteq \Pi_{\mathbb{A}^+}(R)$,*
- *$R\restriction_{L_m}$, the restriction of R to L_m, admits a uniformisation in $\mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$,*

then $R\restriction_{L_m}$ admits a uniformisation obtained as a union of $\langle\binom{\mathbb{A}}{\mathbb{B}}, Z\rangle$-maps, where Z is the natural number $(N + |S_R| + 1) \times \big(|\mathbb{A}| \times (|S_R| + 1)\big)^{|\mathbb{A}_{\trianglerighteq N}(m)|}$, S_R being the syntactic semigroup of R.

We recall that $\mathbb{A}_{\trianglerighteq N}(m)$ is the set of letters of $\mathbb{A}$ that are mapped to $\trianglerighteq_N$ by m.

Proof. The lemma is shown by induction on $|\mathbb{A}_{\trianglerighteq N}(m)|$.

First, we study the case when $|\mathbb{A}_{\trianglerighteq N}(m)| = 0$, which means that m maps every letter a of $\mathbb{A}$ to a natural number smaller than N. Let $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ be a regular relation, and let us suppose that $L_m \subseteq \Pi_{\mathbb{A}^+}(R)$ and that $R\restriction_{L_m}$ admits a uniformisation $F \in \mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$: according to

Proposition 2.17, F can be written as $\bigsqcup_{f\in\Upsilon} F_f$, where Υ is a set of $\langle\binom{\mathbb{A}}{\mathbb{B}}, M\rangle$-maps, M being some natural number.

Let $f \in \Upsilon$, and let $\binom{w}{\sigma}$ be a pair in F_f. Since $w \in L_m$, all letters of $\mathbb{A}$ have at most $N{-}1$ occurrences in w, by assumption on m. Hence, each pair $\binom{a}{b} \in \binom{\mathbb{A}}{\mathbb{B}}$ has at most $N{-}1$ occurrences in $\binom{w}{\sigma}$. Therefore, each $f \in \Upsilon$ is in particular a $\langle\binom{\mathbb{A}}{\mathbb{B}}, N\rangle$-map, and we have the result: $R{\restriction}_{L_m}$ admits a uniformisation in $\mathbf{M}_N[\binom{\mathbb{A}}{\mathbb{B}}]$. Since $N \leq N + |S_R| + 1$, which is the overall number Z of our statement here, we can conclude this case $|\mathbb{A}_{\trianglerighteq N(d)}| = 0$.

Now, we suppose that q is a positive natural number such that the lemma holds for all N and m such that $|\mathbb{A}_{\trianglerighteq N}(m)| < q$, and we want to show that the lemma still holds for all N and m such that $|\mathbb{A}_{\trianglerighteq N}(m)| = q$.

Let N be a natural number, and let m be an $\langle\mathbb{A}, N\rangle$-map such that $|\mathbb{A}_{\trianglerighteq N}(m)| = q$.

For the sake of simplicity, we assume without loss of generality that $a_0, \ldots, a_{p-1}$ are the letters in $\mathbb{A}_N(m)$, and $a'_0, \ldots a'_{q-1}$ are the letters in $\mathbb{A}_{\trianglerighteq N}(m)$: m maps each a_k to some natural number smaller than N, and each a'_k to $\trianglerighteq_N$.

Let us suppose that the assumptions of the lemma hold, meaning that $L_m \subseteq \Pi_{\mathbb{A}^+}(R)$ and that $R{\restriction}_{L_m}$ admits a uniformisation F in $\mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$: again by Proposition 2.17, F can be written as $\bigsqcup_{f\in\Upsilon} F_f$, where Υ is a set of $\langle\binom{\mathbb{A}}{\mathbb{B}}, M\rangle$-maps, M being some natural number. If $M < N$, then we have our result. Let us suppose that $M \geq N$.

For each $f \in \Upsilon$, since F_f is functional, we know from Lemma 2.27 that for all $k \in p$, there exists at most one $b \in \mathbb{B}$ such that $f(\binom{a_k}{b}) \neq 0$. We denote this b by $b_{k,f}$. If $f(\binom{a_k}{b}) = 0$ for all $b \in \mathbb{B}$, then we define $b_{k,f}$ as being any $b \in \mathbb{B}$. Let $k \in p$. Since $m(a_k) < N$, and $\Pi_{\mathbb{A}^+}(F_f) \subseteq L_m$, $f(\binom{a_k}{b_{k,f}})$ is necessarily equal to $m(a_k)$.

Also by Lemma 2.27, there exists, for every $k \in q$, exactly $b' \in \mathbb{B}$ such that $f(\binom{a'_k}{b'}) \neq 0$, and we denote it by $b'_{k,f}$. Here, $f(\binom{a'_k}{b'_{k,f}})$ may have different value: it can be either $\trianglerighteq_M$, or any natural number in $\{N, N{+}1, \ldots, M{-}1\}$.

Let us show that among these f's in Υ, there must exist one that is *maximal* in the sense that f maps $\binom{a'_k}{b'_{k,f}}$ to $\trianglerighteq_M$, for each $k \in q$. To show this, it is enough to consider the word $w = a_0^{m(a_0)} \cdot a_1^{m(a_1)} \cdots a_{p-1}^{m(a_{p-1})} \cdot {a'_0}^M \cdots {a'_{q-1}}^M$: it has exactly $m(a_k)$ occurrences of each a_k, and M occurrences of each a'_k. Since $M \geq N$, w is in $L_m \subseteq \Pi_{\mathbb{A}^+}(R) = \Pi_{\mathbb{A}^+}(\bigsqcup_{f\in\Upsilon} F_f) = \bigsqcup_{f\in\Upsilon} \Pi_{\mathbb{A}^+}(F_f)$, and therefore there exists some $f \in \Upsilon$ such that w is in $\Pi_{\mathbb{A}^+}(f)$. This f is necessarily maximal: for all $k \in q$, $f(\binom{a'_k}{b'_{k,f}})$ cannot be a natural number smaller than M (since w is in the projection and $|w|_{a'_k} = M$), which means that f necessarily maps $\binom{a'_k}{b'_{k,f}}$ to $\trianglerighteq_M$.

Now that we have our maximal $f \in \Upsilon$, we define P as the natural number $\big((N \dot{-} 1)\times|\mathbb{A}|+$

$1\big) \times (|S_R|+1)$, where $N \dot{-} 1$ denotes $N-1$ if $N \geq 1$, and 0 if $N = 0$. Notice that $P \geq N$. We define f' as the $\big\langle \big(\begin{smallmatrix}\mathbb{A}\\ \mathbb{B}\end{smallmatrix}\big), P\big\rangle$-map induced from f: it maps each pair $\big(\begin{smallmatrix}a_k\\ b_{k,f}\end{smallmatrix}\big)$ to $m(a_k) < N \leq P$, each pair $\big(\begin{smallmatrix}a'_k\\ b'_{k,f}\end{smallmatrix}\big)$ to $\trianglerighteq_P$, and all other pairs to 0. The obtained relation $F_{f'}$ is functional, and its projection onto $\mathbb{A}^+$ is included in L_m. We prove now that the relation is included in R (and therefore in $R{\upharpoonright}_{L_m}$).

Claim 2.29. $F_{f'} \subseteq R{\upharpoonright}_{L_m}$.

Proof. We consider some pair $\big(\begin{smallmatrix}w\\ \sigma\end{smallmatrix}\big)$ in $F_{f'}$, in order to prove that it is also in R. For all $i < |w|$, we define $\gamma_w(i)$ as being the element $h_{S_R}(\big(\begin{smallmatrix}w(0)\cdots w(i)\\ \sigma(0)\cdots\sigma(i)\end{smallmatrix}\big))$, in S_R. The function γ_w is therefore a finite word over the alphabet S_R and of length $|\gamma_w| = |w|$. Notice that for all $0 \leq i < j < |w|$, we have $\gamma_w(j) = \gamma_w(i) * h_{S_R}(\big(\begin{smallmatrix}w(i+1)\cdots w(j)\\ \sigma(i+1)\cdots\sigma(j)\end{smallmatrix}\big))$.

We define that a convex subset X of $\mathrm{Dom}(w)$ is *maximal* if each position of X is labelled by one of the a'_k's, and if for every convex subset Y of $\mathrm{Dom}(w)$ with $X \subset Y$, there exists some position of $Y \setminus X$ labelled by one of the a_k's. Figure 2.9 depicts an example of such maximal convex subset. The factor $w{\upharpoonright}_X$ of w is called *maximal* if the subset X is.

			X							
w:	$\cdots$	a_3	a'_0	a'_7	a'_6	$\cdots$	a'_3	a'_2	a_4	$\cdots$
σ:	$\cdots$	b_3	b'_0	b'_7	b'_6	$\cdots$	b'_3	b'_2	b_4	$\cdots$

Figure 2.9: This subset X of $\mathrm{Dom}(w)$ is maximal.

Claim 2.30. *The word w admits at most $(N \dot{-} 1) \times |\mathbb{A}| + 1$ maximal factors.*

Proof. Since $w \in L_m$, it admits, for each $k \in p$, at most $N \dot{-} 1$ positions labelled by a_k, and therefore at most $(N \dot{-} 1) \times |\mathbb{A}|$ positions labelled by one of the a_k's. A maximal factor of w being defined either as the set of positions between two such positions, or as the set of positions before (resp. after) the first (resp. last) such position, which justifies the claim. $\square$

Now, let us consider the positions of w labelled by $\big(\begin{smallmatrix}a'_0\\ b'_{0,f}\end{smallmatrix}\big)$. By the definition of $F_{f'}$, there are at least $P = \big((N \dot{-} 1)\times|\mathbb{A}| + 1\big) \times (|S_R| + 1)$ such positions. Since there are at most $(N \dot{-} 1)\times|\mathbb{A}| + 1$ maximal factors in w, there exists at least one restricted to a convex subset $X \subseteq \mathrm{Dom}(w)$ which contains at least $|S_R| + 1$ occurrences of a'_0. Thus, among these positions in X labelled by a'_0 in w, there must be two, distinct, labelled by the same element of S_R in γ_w: $i < j \in X$ are such that $\binom{w(i)}{\sigma(i)} = \binom{w(j)}{\sigma(j)} = \binom{a'_0}{b'_{0,f}}$ and $\gamma_w(i) = \gamma_w(j)$.

We define w_0 as the finite word $w(0)\cdots w(i)\cdot \big(w(i{+}1)\cdots w(j)\big)^M \cdot w(j{+}1)\cdots w(|w|{-}1)$: it is basically w but with the factor $w(i{+}1)\cdots w(j)$ appearing M times instead of just once. Similarly, we define σ_0 as $\sigma(0)\cdots\sigma(i)\cdot\big(\sigma(i{+}1)\cdots\sigma(j)\big)^M\cdot\sigma(j{+}1)\cdots\sigma(|w|{-}1)$.

A first thing to notice is that the pair $\binom{w_0}{\sigma_0}$ is in R if and only if $\binom{w}{\sigma}$ also is. Indeed, we have:

$$\begin{array}{rcl}
h_{S_R}(\binom{w_0}{\sigma_0}) & = & h_{S_R}(\binom{w(0)\cdots w(i)}{\sigma(0)\cdots\sigma(i)}) * h_{S_R}(\binom{w(i+1)\cdots w(j)}{\sigma(i+1)\cdots\sigma(j)})^M * h_{S_R}(\binom{w(j+1)\cdots w(|w|-1)}{\sigma(j+1)\cdots\sigma(|w|-1)}) \\
& = & \gamma_w(i) * h_{S_R}(\binom{w(i+1)\cdots w(j)}{\sigma(i+1)\cdots\sigma(j)}) * h_{S_R}(\binom{w(i+1)\cdots w(j)}{\sigma(i+1)\cdots\sigma(j)})^{M-1} * h_{S_R}(\binom{w(j+1)\cdots w(|w|-1)}{\sigma(j+1)\cdots\sigma(|w|-1)}) \\
& = & \gamma_w(j) * h_{S_R}(\binom{w(i+1)\cdots w(j)}{\sigma(i+1)\cdots\sigma(j)})^{M-1} * h_{S_R}(\binom{w(j+1)\cdots w(|w|-1)}{\sigma(j+1)\cdots\sigma(|w|-1)}) \\
& = & \gamma_w(i) * h_{S_R}(\binom{w(i+1)\cdots w(j)}{\sigma(i+1)\cdots\sigma(j)})^{M-1} * h_{S_R}(\binom{w(j+1)\cdots w(|w|-1)}{\sigma(j+1)\cdots\sigma(|w|-1)}) \\
& \vdots & \\
& = & \gamma_w(i) * h_{S_R}(\binom{w(i+1)\cdots w(j)}{\sigma(i+1)\cdots\sigma(j)}) * h_{S_R}(\binom{w(j+1)\cdots w(|w|-1)}{\sigma(j+1)\cdots\sigma(|w|-1)}) \\
& = & \gamma_w(|w|-1) \\
& = & h_{S_R}(\binom{w}{\sigma}).
\end{array}$$

Therefore, $\binom{w_0}{\sigma_0} \in h_{S_R}^{-1}(h_{S_R}(R)) = R$ if and only if $\binom{w}{\sigma} \in h_{S_R}^{-1}(h_{S_R}(R)) = R$, and we have our equivalence.

Moreover, since in $\binom{w}{\sigma}$, all positions between i and j are labelled by letters among the $\binom{a'_k}{b'_{k,f}}$'s, we have for each $k \in p$, $|\binom{w_0}{\sigma_0}|_{\binom{a_k}{b_{k,f}}} = |\binom{w}{\sigma}|_{\binom{a_k}{b_{k,f}}} = m(a_k)$, and since $\binom{w(i)}{\sigma(i)} = \binom{w(j)}{\sigma(j)} = \binom{a'_0}{b'_{0,f}}$, the letter $\binom{a'_0}{b'_{0,f}}$ has at least M occurrences in $\binom{w_0}{\sigma_0}$. To summarise, our pair $\binom{w_0}{\sigma_0}$ is such that:

- $|\binom{w_0}{\sigma_0}|_{\binom{a_k}{b_{k,f}}} = m(a_k)$ for each $k \in p$,
- $|\binom{w_0}{\sigma_0}|_{\binom{a'_0}{b'_{0,f}}} \geq M$,
- $\binom{w_0}{\sigma_0} \in R$ if and only if $\binom{w}{\sigma} \in R$.

Now, we can repeat the same process, but on w_0, and focusing on the occurrences of $\binom{a'_1}{b'_{1,f}}$, we obtain a pair $\binom{w_1}{\sigma_1}$ which is such that:

- $|\binom{w_1}{\sigma_1}|_{\binom{a_k}{b_{k,f}}} = m(a_k)$ for each $k \in p$,
- $|\binom{w_1}{\sigma_1}|_{\binom{a'_0}{b'_{0,f}}} \geq M$,
- $|\binom{w_1}{\sigma_1}|_{\binom{a'_1}{b'_{1,f}}} \geq M$,
- $\binom{w_1}{\sigma_1} \in R$ if and only if $\binom{w}{\sigma} \in R$.

When we inductively repeat the procedure, for each of the $\binom{a'_k}{b'_{k,f}}$'s, we obtain, after q steps, a pair $\binom{w_{q-1}}{\sigma_{q-1}}$ that satisfies:

– $|\binom{w_{q-1}}{\sigma_{q-1}}|_{\binom{a_k}{b_{k,f}}} = m(a_k)$ for each $k \in p$,

– $|\binom{w_{q-1}}{\sigma_{q-1}}|_{\binom{a'_k}{b'_{k,f}}} \geq M$ for each $k \in q$,

– $\binom{w_{q-1}}{\sigma_{q-1}} \in R$ if and only if $\binom{w}{\sigma} \in R$.

By the definition of F_f, the pair $\binom{w_{q-1}}{\sigma_{q-1}}$ is in F_f and therefore in R. Thus, $\binom{w}{\sigma}$ is in R, and we have proven our inclusion: $F_{f'} \subseteq R\restriction_{L_m}$. □

As we defined f' from f, we define now m' from m: m' is the $\langle \mathbb{A}, P\rangle$-map that maps each a_k to $m(a_k)$ and each a'_k to $\trianglerighteq_P$. The language $L_{m'}$, included in L_m, is exactly the projection of $F_{f'}$ onto $\mathbb{A}^+$, and therefore, the relation $F_{f'}$, which we proved in Claim 2.29 to be included in R, is a uniformisation of $R\restriction_{m'}$.

It remains to take care of the restriction of R to the set $L_m \setminus L_{m'}$, which we denote by L. We use the induction hypothesis for this. First, L is in $\mathbf{M}_N[\mathbb{A}]$ by definition, and therefore in $\mathbf{M}_P[\mathbb{A}]$, hence, L, as a union of languages in $\mathbf{M}_P[\mathbb{A}]$, is in $\mathbf{M}_P[\mathbb{A}]$, by Corollary 2.18. Thus, we write L as $\bigsqcup_{o\in\Lambda} L_o$, with Λ being a set of $\langle \mathbb{A}, P\rangle$-maps. Our point is to prove that for each $o \in \Lambda$, $R\restriction_{L_o}$ admits a uniformisation that is the union of $\left\langle \binom{\mathbb{A}}{\mathbb{B}}, (P + |S_R| + 1) \times \big(|\mathbb{A}| \times (|S_R| + 1)\big)^q \right\rangle$-maps.

Let $o \in \Lambda$. Because $L_o \subseteq L_m$, we have $o(a_k) = m(a_k)$ for each $k \in p$. Also, it is not possible for each $o(a'_k)$ to be equal to $\trianglerighteq_P$, for o would be exactly the function m'. Therefore, we have $|\mathbb{A}_{\trianglerighteq_P}(o)| < |\mathbb{A}_{\trianglerighteq_P}(m')| = q$.

We have $L_o \subseteq L_m \subseteq R$. Hence, to apply our inductive hypothesis on the pair $\langle o, P\rangle$, we have to prove that $R\restriction_{L_o}$ admits a uniformisation in $\mathbf{FO}[\,]$, and this is immediate: $\bigsqcup_{f\in\Upsilon} F_f$ is a uniformisation of $R\restriction_{L_m}$ in $\mathbf{FO}[\,]$, and therefore $\binom{L_o}{\mathbb{B}^+} \cap \bigsqcup_{f\in\Upsilon} F_f$ is a uniformisation of $R\restriction_{L_o}$ in $\mathbf{FO}[\,]$. Indeed, $\binom{L_o}{\mathbb{B}^+}$ is clearly definable in $\mathrm{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$, and we know that this class is closed under Boolean combinations.

Since $|\mathbb{A}_{\trianglerighteq_P}(o)| < q$, we can apply our induction hypothesis: R restricted to L_o admits a uniformisation that is the union of $\left\langle \binom{\mathbb{A}}{\mathbb{B}}, (P + |S_R| + 1) \times \big(|\mathbb{A}| \times (|S_R| + 1)\big)^{q-1} \right\rangle$-maps. Since $P \leq \big(N \times |\mathbb{A}| + 1\big) \times (|S_R| + 1)$, we also have:

$$(P + |S_R| + 1) \times \big(|\mathbb{A}| \times (|S_R| + 1)\big)^{q-1} \leq (N + |S_R| + 1) \times \big(|\mathbb{A}| \times (|S_R| + 1)\big)^{q} = Z,$$

and therefore $R\restriction_{L_o}$ admits a uniformisation F_o that is the union of $\left\langle \binom{\mathbb{A}}{\mathbb{B}}, Z \right\rangle$-maps.

Finally, we can conclude: R restricted to L_m admits $F_{f'} \sqcup \bigsqcup_{o\in\Lambda} F_o$ as a uniformisation that is a union of $\left\langle \binom{\mathbb{A}}{\mathbb{B}}, (N + |S_R| + 1) \times \big(|\mathbb{A}| \times (|S_R| + 1)\big)^q \right\rangle$-maps, and we have proven the lemma. $\square$

We directly deduce the corollary, stating about the whole relation R:

Corollary 2.31. *Let $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ be a regular relation. We suppose that its projection $\Pi_{\mathbb{A}^+}(R)$ belongs to $\mathbf{M}_N[\mathbb{A}]$ for some $N \in \mathbb{N}$. Then, if R admits a uniformisation in $\mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$, it admits a uniformisation in $\mathbf{M}_Z[\binom{\mathbb{A}}{\mathbb{B}}]$, with Z being the natural number $(N + |S_R| + 1) \times (|\mathbb{A}| \times (|S_R| + 1))^{|\mathbb{A}|}$.*

Proof. We can write $\Pi_{\mathbb{A}^+}(R)$ as $\bigsqcup_{m\in\Lambda} L_m$, with Λ being a set of $\langle\mathbb{A}, N\rangle$-maps. Let us suppose that R admits indeed a uniformisation F in $\mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$. Then for each $m \in \Lambda$, $R{\restriction}_{L_m}$ admits $F{\restriction}_{L_m} = F \cap \binom{L_m}{\mathbb{B}^+}$ as uniformisation in $\mathbf{FO}[\;]$. Moreover, by definition, each L_m is included in $\Pi_{\mathbb{A}^+}(R)$. Therefore, by Lemma 2.28, we know that $R{\restriction}_{L_m}$ admits a uniformisation F_m in $\mathbf{M}_Z[\binom{\mathbb{A}}{\mathbb{B}}]$. Hence, $\bigsqcup_{m\in\Lambda} F_m$, the union of all of these, is a uniformisation of R in $\mathbf{M}_Z[\binom{\mathbb{A}}{\mathbb{B}}]$. $\square$

We can deduce the main theorem: it is decidable whether a given regular relation admits a uniformisation in the class $\mathbf{FO}[\;]$.

Theorem 2.32. *There exists a algorithm that inputs an* $\mathrm{MSO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$ *sentence φ, $\mathbb{A}$ and $\mathbb{B}$ being any alphabets, and outputs an* $\mathrm{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$ *formula ψ such that the relation $\mathcal{L}(\psi)$ uniformises the relation $\mathcal{L}(\varphi)$ if there exists one, and* **NO** *in the opposite case.*

Proof. Let $R = \mathcal{L}(\varphi) \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$.

Proposition 1.47 in Chapter 1 gives us a formula φ_{proj} that defines $\Pi_{\mathbb{A}^+}(R)$. Inputting φ^π to the algorithm of Corollary 2.23, we either get a **NO**, telling us that $\Pi_{\mathbb{A}^+}(R)$ is not in $\mathbf{FO}[\;]$, or a set Λ of $\langle\mathbb{A}, N\rangle$-maps, N being some natural number, such that $\Pi_{\mathbb{A}^+}(R) = \bigsqcup_{m\in\Lambda} L_m$.

In the former case, we output **NO**, since a relation whose projection is not in $\mathbf{FO}[\mathbb{A}]$ cannot have a uniformisation in $\mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$, by Corollary 2.25.

In the latter case, we can use Corollary 2.31, with the natural N: we know that if R admits a uniformisation in $\mathbf{FO}[\;]$, then it must admit one of the shape $\bigsqcup_{f\in\Upsilon} F_f$, with Υ being a set of $\mathbf{M}_Z[\binom{\mathbb{A}}{\mathbb{B}}]$-maps, with Z being the natural number $(N+|S_R|+1)\times(|\mathbb{A}|\times(|S_R|+1))^{|\mathbb{A}|}$ (which we can compute, since we can compute S_R—see Proposition 1.15 on page 42).

Given a set Υ of $\langle\binom{\mathbb{A}}{\mathbb{B}}, Z\rangle$-maps, we can test if $F_\Upsilon = \bigsqcup_{f\in\Upsilon} F_f$ is a uniformisation of R. Indeed, remember that F_Υ is a uniformisation of R if the following three conditions hold:

i) F_Υ is functional,

ii) $F_\Upsilon \subseteq R$,

iii) $\Pi_{\mathbb{A}^+}(F) = \Pi_{\mathbb{A}^+}(R)$.

First, we can ensure Condition i) by considering only the sets Υ whose maps f induce functions, meaning such that for all $a \in \mathbb{A}$, there exists at most one $b \in \mathbb{B}$ such that $f(\binom{a}{b}) \neq 0$ (see Lemma 2.27). Then, we can test the inclusion $F_\Upsilon \subseteq R$ using the algorithm of Corollary 1.17, since Proposition 2.17 gives us a formula for F_Υ. Finally, we can compute $\Pi_{\mathbb{A}^+}(F_\Upsilon)$ and compare it with $\Pi_{\mathbb{A}^+}(R)$.

Hence, if among the finite amount of sets of $\langle\binom{\mathbb{A}}{\mathbb{B}}, Z\rangle$-maps, we find some whose induced language is a uniformisation of R, we output it (or an $\mathrm{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$ sentence defining it). If not, then we output **NO**, and we are done. $\square$

2.5 Conclusion and further interrogations

In this chapter, we proved both negative and positive results about uniformising relations of finite words in First-Order Logic and its different fragments, which we introduced in Subsection 1.2.4.

First, we proved that most of these fragments, unlike the full class $\mathbf{MSO}[<]$ of regular languages, are too weak to uniformise not only themselves, but also weaker formalisms, as state Propositions 2.1, 2.2, and 2.3. The counterexamples we provided, and the proofs relying on the algebraic characterisations of these classes, give us an idea about what these formalisms are missing in order to construct uniformisations.

Yet, using a characterisation of $\mathbf{FO}[\,]$, which we proved in Section 2.2, we were able to prove two positive results. The first one, stating that $\mathbf{FO}[<]$ is actually strong enough to uniformise relations in $\mathbf{FO}[\,]$, contrasts with our first three negative results, and raises different questions.

First, one could ask whether the class $\mathbf{FO}[<]$ is capable of uniformising more than relations expressible in $\mathbf{FO}[\,]$. For instance, what about the possibility to uniformise in $\mathbf{FO}[<]$ the relations definable in $\mathrm{FO}^2[<]$, the fragment using only two different variables?

Question 2.33. *Does the class* $\mathbf{FO}[<]$ *uniformise the class* $\mathbf{FO}^2[<]$*?*

Indeed, it is noticeable that the class $\mathbf{FO}^2[<]$ contains $\mathbf{FO}[\,]$, for the property of containing at least k occurrences of some letter a can be defined by the $\mathrm{FO}^2[<]$ formula $\exists x_0.\ a(x_0) \wedge \exists x_1{>}x_0.\ a(x_1) \wedge \exists x_0{>}x_1.\ a(x_0) \wedge \ldots$, where k existential quantifiers are involved, and, thus, we can define in $\mathrm{FO}^2[<]$ the languages induced from maps, introduced in Section 2.2. Yet, this inclusion is not directly inherited from the syntax, which explains why we did not picture it on Figure 1.1 on page 40.

The author was not able yet to construct a relation definable in $\mathrm{FO}^2[<]$ that does not admit uniformisation in $\mathbf{FO}[<]$. If, on the other hand, the answer to Question 2.33 is positive, it shall be proven using the semantic characterisation of $\mathbf{FO}^2[<]$ [WI07], as we use the characterisation of $\mathbf{FO}[\,]$ in order to prove Proposition 2.5.

Also, although the chapter focuses on fragments of First-Order Logic, it seems legitimate to study the possibility to uniformise relations in formalisms lying between $\mathrm{FO}[<]$ and $\mathrm{MSO}[<]$. As an example, what about the possibility to uniformise relations in the formalism $\mathrm{FO}[<, \mathrm{mod}]$, which enriches $\mathrm{FO}[<]$ with modular predicate (meaning that it can test if a position is dividable by some natural number [CPS06])?

Another natural question that arises from these results is the existence of a robust non-empty subclass of $\mathbf{MSO}[<]$ that has the possibility to uniformise its relations. For instance, if we consider, for a binary relation $R \in \mathbf{FO}[\,]$, a uniformisation F_R of it in the class $\mathbf{FO}[<]$, then the class $\mathbf{U} := \mathbf{FO}[\,] \cup \{F_R \mid R \in \mathbf{FO}[\,]\} \subset \mathbf{FO}[<] \subset \mathbf{MSO}[<]$ easily satisfies the uniformisation property, but this class is of limited interest: it does not admit any natural closure properties.

In the next chapter, we answer this question negatively, when considering the robustness of varieties of languages, defined in Subsection 1.3.2: any non-empty variety of languages that does not contain all the regular languages necessarily admits a relation that it cannot uniformise.

Another positive result which we obtained in this chapter is on the decidability of uniformising regular relations in $\mathrm{FO}[\,]$. The proof that we proposed rely on both semantic and algebraic characterisations of this formalism, but also on a very strong property of it, which is its closure under projections: if $R \in \mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}]$, then $\Pi_{\mathbb{A}^+}(R) \in \mathbf{FO}[\mathbb{A}]$ (see Corollary 2.21).

This property is very rare among formalisms: natural classes such as $\mathbf{FO}[<]$ or $\mathbf{FO}[\mathsf{s}]$ do not share it. Hence, generalising our proof to other classes, such as these two, does not seem feasible. Yet, we have the following proposition, stating that, even though $\mathbf{FO}^2[<]$ is not closed under projections, it does satisfy a similar property, which we prove in Appendix B:

Proposition 2.34. *Let $\mathbb{A}$ and $\mathbb{B}$ be two alphabets, and let $F \in \mathbf{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, <]$. If F is functional, then its projection $\Pi_{\mathbb{A}^+}(F)$ is in $\mathbf{FO}^2[\mathbb{A}, <]$.*

In particular, if some relation $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ admits a uniformisation F in $\mathbf{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, <]$, then its projection $\Pi_{\mathbb{A}^+}(R)$ must be in $\mathbf{FO}^2[\mathbb{A}, <]$, since it is equal to the projection of F. For this reason, $\mathbf{FO}^2[<]$ would be a natural candidate to see if our proof can somehow be adapted. Yet, at the moment, the author was not able to come with an algorithm deciding the uniformisability in this class.

Chapter 3

$\mathbf{MSO}[<]$ as the unique non-empty variety with the uniformisation property

In the previous chapter, we studied the capacity of certain fragments of First-Order Logic, over finite words, to uniformise each other. We were able to find many examples in which said fragments are unable to provide uniformisations for other classes, even much weaker ones. Now, we focus on the capacity of a class to uniformise its own relations. Indeed, the results of Chapter 2 do not exclude, *a priori*, the existence of a robust self-uniformising class, for example between $\mathbf{FO}[\,]$ and $\mathbf{FO}[<]$.

Recall that a class $\mathbf{C}$ of languages satisfies the uniformisation property if for each relation R in $\mathbf{C}$, there exists a relation F also in $\mathbf{C}$ that is a uniformisation of R. Two very robust classes of regular languages of finite words are known to have this property: the empty class $\emptyset$, trivially, and the full class of regular languages, $\mathbf{MSO}[<]$, by Theorem 1.48. In this chapter, we show that, among varieties of languages (see the definition on page 45), these two classes are the only ones, as states the following theorem, originally proven in [LMS19]:

Theorem 3.1. $\mathbf{MSO}[<]$ *is the unique non-empty variety of languages that satisfies the uniformisation property.*

A graphical representation of Theorem 3.1 is depicted in Figure 3.1.

In the right-hand side ellipse, one can find all the non-empty classes of regular languages that satisfy the convenient closure properties of varieties of languages: the full class $\mathbf{MSO}[<]$ of regular languages, fragments of First-Order Logic, the class $\mathbf{FC}$ of languages that are either finite or *cofinite*[1], the class $\mathbf{EF}$ of empty and full languages introduced on page 55, ...

[1] A language $L \subseteq \mathbb{A}^+$ is called *cofinite* if its complement language $L^c := \mathbb{A}^+ \setminus L$ is finite. The class $\mathbf{FC}$ is naturally closed under Boolean operations, by definition. If $u \in \mathbb{A}^+$, then for any language L, $w \mapsto u{\cdot}w$ and $w \mapsto w{\cdot}u$ are injections from $u^{-1}{\cdot}L$ and $L{\cdot}u^{-1}$, respectively, to L, therefore $\mathbf{FC}$ is also closed under quotients. Finally, if $h\colon \mathbb{A}^+ \to \mathbb{B}^+$ is a homomorphism, then any word $w \in \mathbb{B}^+$ admits at most $|\mathbb{A}| \times |w|$ preimages under h. By this argument, $\mathbf{FC}$ is also closed under preimages of homomorphisms, and we can conclude that it is a variety of languages.

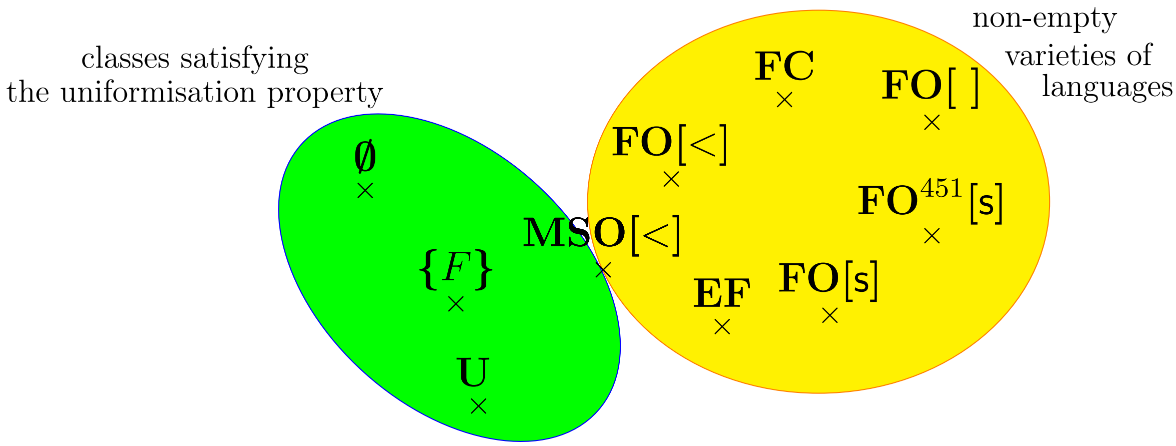


Figure 3.1: A graphical representation of Theorem 3.1.

In the left-hand side ellipse are depicted the classes that can uniformise their relations: one can see **MSO**[<], the empty class $\emptyset$, and some more artificial classes. For instance, a class containing a unique relation F that happens to be functional clearly satisfies the uniformisation property, but it is not a variety of languages since it is not even closed under Boolean operations. On the figure, one can also see **U**, the class we defined on page 87 and that contains relations in **FO**[ ], as well as uniformisations of these in **FO**[<].

Theorem 3.1 tells us that these two ellipses intersect each other exactly in one class: the class **MSO**[<] of regular languages. In other words, the capacity of providing uniformisations characterises **MSO** among all the non-empty varieties of finite semigroups.

Except for its last section, the whole chapter is dedicated to the proof of the theorem. We consider a class **C** of regular languages of finite words, and we suppose that it satisfies the three hypotheses of the theorem (it is non-empty, it is a variety of languages, and it uniformises itself), and we prove that it is necessarily the class **MSO**[<]. The strategy is to depict step by step how **C** must contain more and more regular properties, until we ultimately show that it contains every single regular language.

We begin by proving in Section 3.1 that **C** can test letters: if a is a letter of an alphabet $\mathbb{A}$, then the language $\{w \in \mathbb{A}^+ \mid \text{at least one position of } w \text{ is labelled by } a\}$ is a language of **C**. This is equivalent to saying that **C** contains $\mathbf{FO}^1[\,]$, the class of languages definable by a formula using only one variable and whose signature is composed only of letter tests (in addition to the equality). Indeed, we recall that we gave a semantic characterisation of this class in Proposition 2.17. In Section 3.2, we show that **C** is capable not only of testing its letters, but also of counting them: for any $n \in \mathbb{N}$, also the language $\{w \in \mathbb{A}^+ \mid$

at least n positions of w are labelled by $a\}$ is in $\mathbf{C}$. This time, this capacity is equivalent to $\mathbf{FO}[\,] \subseteq \mathbf{C}$ (again, see Proposition 2.17 for the semantic characterisation of this class).

In Section 3.3, a more important step is completed as we prove that $\mathbf{C}$ can recognise subwords: if $a_0, a_1, \ldots, a_{n-1}$ are letters of an alphabet $\mathbb{A}$, then the language $\mathbb{A}^* \cdot a_0 \cdot \mathbb{A}^* \cdot a_1 \cdot \mathbb{A}^* \cdots \mathbb{A}^* \cdot a_{n-1} \cdot \mathbb{A}^*$ belongs to $\mathbf{C}$. This sort of languages correspond to the *existential fragment* of FO[<], *i.e.* they can be defined by a formula of the shape $\exists x_0, \ldots,\ x_{n-1}.\ \psi(x_0, \ldots, x_{n-1})$, with $\psi(x_0, \ldots, x_{n-1})$ being a FO[<] formula of quantifier depth 0 (it is a Boolean combination of formulae of the shape $a(x_i)$, $x_i < x_j$, or $x_i = x_j$). We refer to [DGK08, Section 6] for the expressive power of said fragment.

Finally, we reach our aim in Section 3.4, when we prove that the recognition by semigroups can be expressed in $\mathbf{C}$, and therefore that every regular language is in $\mathbf{C}$.

In most of the proofs of this chapter, we make use of the algebraic closures of the class $\mathbf{V} := \mathbf{Synt}(\mathbf{C})$ of (finite) syntactic semigroups of languages in $\mathbf{C}$ (see its definition on page 45). Theorem 1.24 tells us that it is a variety of finite semigroups (again, the definition of a v.f.s. can be found on page 46), and that $\mathbf{C}$ is exactly $\mathbf{Reco}(\mathbf{V})$, the class of languages recognised by some semigroup in $\mathbf{V}$.

3.1 Testing letters and extending alphabets

This section is devoted to a first step of the proof of Theorem 3.1: we show that $\mathbf{C}$ must be able to detect which letters appear in the given word. More formally, the main result of this section is the following lemma, and its immediate corollaries:

Lemma 3.2. *For all alphabets $\mathbb{A}_1 \subseteq \mathbb{A}_2$, the language $\{w \in \mathbb{A}_2^+ \mid w(x) \in \mathbb{A}_1$ for each $x \in |w|\}$ is in $\mathbf{C}$.*

In this section, we write $\langle \mathbb{A}_1^+, \mathbb{A}_2 \rangle$ for the language of the lemma (see the remark on page 24). The information about the alphabet $\mathbb{A}_2$ is indeed important: a^+ is a crucially different object whether it is considered as a language over the alphabet $\{a\}$ or over a bigger alphabet $\{a, b\}$.

Before proving the main lemma, we prove a shorter one, stating that $\mathbf{C}$ contains all the full languages:

Lemma 3.3. *For any alphabet $\mathbb{A}$, $\mathbf{C}$ contains the full language $\langle \mathbb{A}^+, \mathbb{A} \rangle$.*

Proof. Since $\mathbf{C}$ is not empty by assumption, it contains some language L over an alphabet $\mathbb{A}_1$. Let S be a semigroup in $\mathbf{V}$ that recognises L. We know that there exists one, since the syntactic semigroup of L is in $\mathbf{V}$. Then, we can consider any homomorphism h from $\mathbb{A}$ to S: we know by Proposition 1.18 on page 43 that there exists some idempotent element e in S, and the function mapping each word w of $\mathbb{A}^+$ to e is a homomorphism. We obtain that $\langle \mathbb{A}^+, \mathbb{A} \rangle = h^{-1}(S)$. Therefore, $\langle \mathbb{A}^+, \mathbb{A} \rangle$ is recognised by S, and therefore it is in $\mathbf{C}$, since $\mathbf{C} = \mathbf{Reco}(\mathbf{V})$. □

From this, we know that $\mathbf{C}$ contain $\mathbf{EF}$, the class of empty and full languages, defined on page 55. Now, we can prove the wanted result:

Proof of Lemma 3.2. To prove it, we show that the semigroup $2 = \{0, 1\}$, with the maximum law max (defined by $\max(0, 0) = 0$ and $\max(0, 1) = \max(1, 0) = \max(1, 1) = 1$), belongs to $\mathbf{V}$. Indeed, if this semigroup is in $\mathbf{V}$, then we consider the homomorphism h from $\mathbb{A}_2^+$ to 2, defined by $h(a) = 0$ for $a \in \mathbb{A}_1$ and $h(a) = 1$ for $a \in \mathbb{A}_2 \setminus \mathbb{A}_1$. Then $\langle \mathbb{A}_1^+, \mathbb{A}_2 \rangle = h^{-1}(\{0\})$ and therefore it belongs to $\mathbf{C}$.

Let R be the full relation $\left\langle \binom{\mathbb{A}}{\mathbb{B}}^+, \binom{\mathbb{A}}{\mathbb{B}} \right\rangle$ between words over $\mathbb{A} = \{x\}$ and words over $\mathbb{B} = \{\square, \triangle\}$. Lemma 3.3 tells us that R is in $\mathbf{C}$, as a full language. By the assumption, $\mathbf{C}$ must contain a uniformisation F of R that is recognised by some semigroup $S \in \mathbf{V}$ via a homomorphism h from $\binom{\mathbb{A}}{\mathbb{B}}^+$ to S: $F = h^{-1}(H)$ with $H \subseteq S$.

Let $N = N(2)$ be the number obtained from Theorem 1.29 on page 49 applied for S and h in the particular case $n = 2$. Consider the word $w = x^N$, and take the unique word σ in $\{\square, \triangle\}^N$ such that $\binom{w}{\sigma} \in F$. For convenience, for all $i < j \leq N$, we write $w_{i,j}$ (resp. $\sigma_{i,j}$) for the word $w(i) \cdots w(j-1)$ (resp. $\sigma(i) \cdots \sigma(j-1)$), and $s_{i,j}$ for $h(\binom{w_{i,j}}{\sigma_{i,j}})$.

By the definition of N, we know that there exists an idempotent element e of S, and numbers $i < j < k \leq N$, such that $s_{i,j} = s_{j,k} = e$. Since $j - i > 0$ and $|\mathbb{B}| \geq 2$, we can define a word $\sigma'_{i,j} \in \mathbb{B}^{j-i}$ distinct from $\sigma_{i,j}$ (typically: by changing all □'s of $\sigma_{i,j}$ into △'s, and vice-versa). We define now $s'_{i,j} = h(\binom{w_{i,j}}{\sigma'_{i,j}})$. An illustration of this paragraph is depicted on Figure 3.2.

As e is an idempotent of S, for all $\ell \geq 1$ we have $s_{0,i} \cdot e^\ell \cdot s_{k,N} = s_{0,N} \in H$. Recall that $\sharp$ is a natural number such that for every $s \in S$ the element $s^\sharp$ is idempotent (once again, see Proposition 1.18 on Page 43). Considering the particular case of the above equality for ℓ being $3 \times \sharp$, we obtain:

$$\binom{w_{0,i}}{\sigma_{0,i}} \cdot \left(\binom{w_{i,j}}{\sigma_{i,j}} \cdot \binom{w_{i,j}}{\sigma_{i,j}} \cdot \binom{w_{i,j}}{\sigma_{i,j}} \right)^\sharp \cdot \binom{w_{k,N}}{\sigma_{k,N}} \in F.$$

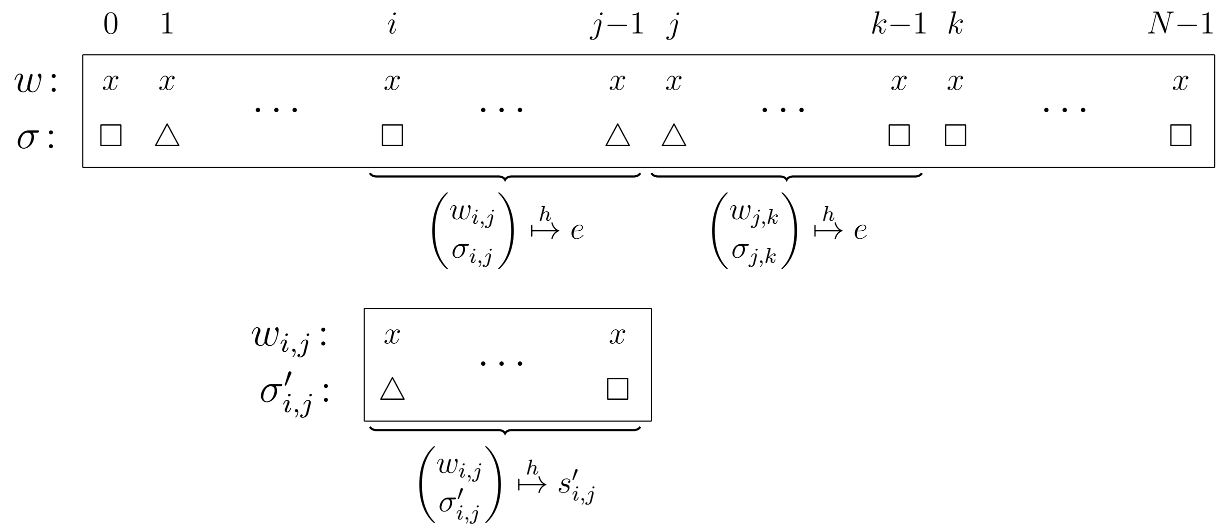


Figure 3.2: Ramsey's theorem applied to the pair $\langle w, \sigma \rangle$, and the definition of $s'_{i,j}$.

As $\sigma_{i,j} \neq \sigma'_{i,j}$ and F is a uniformisation, we know that:

$$\binom{w_{0,i}}{\sigma_{0,i}} \cdot \left(\binom{w_{i,j}}{\sigma_{i,j}} \cdot \binom{w_{i,j}}{\sigma'_{i,j}} \cdot \binom{w_{i,j}}{\sigma_{i,j}} \right)^{\sharp} \cdot \binom{w_{k,N}}{\sigma_{k,N}} \notin F.$$

This implies that the element e' defined as $\left(s_{i,j} \cdot s'_{i,j} \cdot s_{i,j}\right)^{\sharp}$ is different from e. Now, we set $s_0 = e$ and $s_1 = e'$. As both e and e' are idempotents of S, we have $s_0 * s_0 = s_0$ and $s_1 * s_1 = s_1$. But we also have $s_0 * s_1 = s_1$. Indeed, if $\sharp = 1$, then $s_0 * s_1 = e * e * s'_{i,j} * s_{i,j} = e * s'_{i,j} * s_{i,j} = s_1$, and if $\sharp > 1$, then $s_0 * s_1 = e * (e * s'_{i,j} * s_{i,j})^{\sharp} = e * (e * s'_{i,j} * s_{i,j}) * (e * s'_{i,j} * s_{i,j})^{\sharp - 1} = (e * s'_{i,j} * s_{i,j}) * (e \cdot s'_{i,j} * s_{i,j})^{\sharp-1} = (e * s'_{i,j} * s_{i,j})^{\sharp} = s_1$. By the same argument, we have $s_1 * s_0 = s_1$. Therefore, the subset $\{s_0, s_1\}$ of S is a subsemigroup and it is isomorphic to 2 with the law max. Because $\mathbf{V}$ is closed under subsemigroups and under images under surjective homomorphisms, said semigroup is in $\mathbf{V}$, and this concludes the proof of the lemma. $\square$

We can now state that $\mathbf{C}$ is closed under extensions of alphabets, and, even more, it recognises languages independently of the alphabets:

Corollary 3.4. *Let $\mathbb{A}_1 \subseteq \mathbb{A}_2$ be two alphabets and $L \subseteq \mathbb{A}_1^+$, regular. Then $\langle L, \mathbb{A}_1 \rangle \in \mathbf{C}$ if and only if $\langle L, \mathbb{A}_2 \rangle \in \mathbf{C}$.*

Proof. For the forward implication, suppose that $S \in \mathbf{V}$ recognises L via some homomorphism $h_1 \colon \mathbb{A}_1^+ \to S$. We showed in the proof of Lemma 3.2 that the semigroup 2 with the law max is in $\mathbf{V}$. This implies that $S_1 \times 2$, with the natural product law, is also in $\mathbf{V}$ (because $\mathbf{V}$ is closed under products, by the definition of a v.f.s.). Now, let h_2 be the homomorphism from $\mathbb{A}_2^+$ to $S \times 2$ defined by $h_2(a) = \langle h_1(a), 0 \rangle$ for $a \in \mathbb{A}_1$, and $h_2(a) = \langle s, 1 \rangle$ for $a \in \mathbb{A}_2 \setminus \mathbb{A}_1$, where s is any element of S. Then we have clearly $L = h_2^{-1}(h_1(L) \times \{0\})$: $\langle L, \mathbb{A}_2 \rangle$ is recognised by S and therefore it is in $\mathbf{C}$.

For the backward implication, let us suppose that $S \in \mathbf{V}$ recognises L via a homomorphism $h_2 \colon \mathbb{A}_2^+ \to S$. Let h_1 be the homomorphism from $\mathbb{A}_1^+$ to S simply defined by $h_1(w) = h_2(w)$ for $w \in \mathbb{A}_1^+$. By assumption, we have $L = h_2^{-1}(h_2(L)) = h_2^{-1}(h_1(L)) = h_1^{-1}(h_1(L))$: $\langle L, \mathbb{A}_1 \rangle$ is recognised by S and therefore it is in $\mathbf{C}$. $\square$

Thus, we can simply state that a language L is in the class $\mathbf{C}$, without being specific about the alphabet considered. Until the end of the chapter, we will speak simply about a language L, instead of a pair $\langle L, \mathbb{A} \rangle$.

Now, if a is any letter of an alphabet $\mathbb{A}$, we denote by $[\exists a]_{\mathbb{A}}$ the language of finite words over $\mathbb{A}$ that contain at least one occurrence of a, and by $\mathbb{A}^{\oplus}$ we denote the language of finite words over $\mathbb{A}$ in which each letter of $\mathbb{A}$ appears at least once. Using Lemma 3.2, we can deduce that these two languages are in $\mathbf{C}$.

Corollary 3.5. *Let $\mathbb{A}$ be an alphabet and a be any letter of it. Then $[\exists a]_{\mathbb{A}}$ and $\mathbb{A}^{\oplus}$ are both languages of $\mathbf{C}$.*

Proof. It is enough to observe that $[\exists a]_{\mathbb{A}} = \big((\mathbb{A} \setminus \{a\})^+\big)^{\mathrm{c}}$, where the complement is over the full language $\mathbb{A}^+$. Lemma 3.2 tells us that $(\mathbb{A} \setminus \{a\})^+$, as a language over $\mathbb{A}$, is in $\mathbf{C}$, and, since $\mathbf{C}$ is closed under complements, $[\exists a]_{\mathbb{A}} \in \mathbf{C}$.

Now, $\mathbb{A}^{\oplus} = \bigcap_{a \in \mathbb{A}} [\exists a]_{\mathbb{A}}$, and therefore $\mathbb{A}^{\oplus} \in \mathbf{C}$. $\square$

This corollary reveals that $\mathbf{C}$ has at least the expressive power of $\mathbf{FO}^1[\,]$, the one-variable fragment of $\mathbf{FO}$ admitting only letter tests in its signature (see Proposition 2.17 on page 71 in the previous chapter for the expressive power of this class).

3.2 Counting letters

Our second step towards Theorem 3.1 is to prove that $\mathbf{C}$ is able to test single occurrences of letters. Let $\mathbb{A}_1$ and $\mathbb{A}_2$ be two alphabets, with $\mathbb{A}_1 \subseteq \mathbb{A}_1$. We denote by $[\exists^{=1} \mathbb{A}_1]_{\mathbb{A}_2}$ the language of finite words over $\mathbb{A}_2$ that contain exactly one occurrence of a letter of $\mathbb{A}_1$. We prove in this section that this language is in $\mathbf{C}$:

Lemma 3.6. *Let $\mathbb{A}_1 \subseteq \mathbb{A}_2$ be two alphabets, then the language $[\exists^{=1}\mathbb{A}_1]_{\mathbb{A}_2}$ is in* $\mathbf{C}$.

Similarly as for Lemma 3.2, the above lemma can be equivalently expressed by saying that $\mathbf{FO}^2[\,] \subseteq \mathbf{C}$ (again, we expose the expressive power of this class in Proposition 2.17).

Proof. To prove the lemma, consider three distinct letters, x, y, and z, and four distinct symbols $\otimes$, $\oplus$, $\ominus$, and $\odot$. Let R^x and R^y be the relations defined as:

$$R^x = \left\{ \tbinom{x}{\oplus}, \tbinom{x}{\ominus}, \tbinom{y}{\otimes}, \tbinom{z}{\odot} \right\}^{\oplus},$$
$$R^y = \left\{ \tbinom{x}{\otimes}, \tbinom{y}{\oplus}, \tbinom{y}{\ominus}, \tbinom{z}{\odot} \right\}^{\oplus},$$

where $(\cdot)^{\oplus}$ is the operation defined on page 93. These two relations are over the product alphabet $\tbinom{\{x,y,z\}}{\{\oplus,\ominus,\otimes,\odot\}}$.

It is immediate to see that these two relations are disjoint, and we know that they are both in $\mathbf{C}$, by Corollary 3.5. Finally, we define $R = R^x \sqcup R^y$, which is also in $\mathbf{C}$ because the latter is closed under unions.

Since $\mathbf{C}$ satisfies the uniformisation property, there exists $F \in \mathbf{C}$ uniformising R. Let S be a semigroup in $\mathbf{V}$ recognising F: $F = h^{-1}(H)$, where h is a homomorphism from $\tbinom{\{x,y,z\}}{\{\oplus,\ominus,\otimes,\odot\}}^+$ to S and H is a subset of S.

Now, for $p, q \in \mathbb{N}$, we define L^x_p and L^y_q as the following two relations:

$$L^x_p = \left\{ \tbinom{u}{\sigma} \in \left\{ \tbinom{x}{\otimes}, \tbinom{z}{\odot} \right\}^+ \mid \otimes \text{ appears exactly } p \text{ times in } \sigma \right\},$$
$$L^y_q = \left\{ \tbinom{v}{\tau} \in \left\{ \tbinom{y}{\otimes}, \tbinom{z}{\odot} \right\}^+ \mid \otimes \text{ appears exactly } q \text{ times in } \tau \right\}.$$

Once again, all these languages are pairwise disjoint. Notice furthermore that we have the equality $\bigsqcup_{p\in\mathbb{N}} L^x_p = \left\{ \tbinom{x}{\otimes}, \tbinom{z}{\odot} \right\}^+$. Similarly $\bigsqcup_{q\in\mathbb{N}} L^y_q = \left\{ \tbinom{y}{\otimes}, \tbinom{z}{\odot} \right\}^+$.

Claim 3.7. *At least one of the two following propositions is true:*
- *for all $p \geq 2$ we have $h(L^x_1) \cap h(L^x_p) = \emptyset$,*
- *for all $q \geq 2$ we have $h(L^y_1) \cap h(L^x_q) = \emptyset$.*

Proof. Assume the contrary and take:

$$p \geq 2,\ \tbinom{u_1}{\sigma_1} \in L^x_1, \text{ and } \tbinom{u_p}{\sigma_p} \in L^x_p \text{ such that } h(\tbinom{u_1}{\sigma_1}) = h(\tbinom{u_p}{\sigma_p}); \text{ and} \tag{3.1}$$
$$q \geq 2,\ \tbinom{v_1}{\tau_1} \in L^y_1, \text{ and } \tbinom{v_p}{\tau_q} \in L^y_q \text{ such that } h(\tbinom{v_1}{\tau_1}) = h(\tbinom{v_q}{\tau_q}). \tag{3.2}$$

Let w be the word $u_p \cdot v_q \cdot z$. By construction, it is in the projections onto $\{x, y, z\}^+$ of both R^x and R^y, since it admits at least two x's, two y's, and one z. Therefore, it is in $\Pi_{\{x,y,z\}^+}(R)$. Let ν be the unique word over $\{\oplus, \ominus, \otimes, \odot\}$ such that $\binom{w}{\nu} \in F$. Since we have the inequality $F \subseteq R = R^x \sqcup R^y$, we know that $\binom{w}{\nu}$ is either in R^x or in R^y. Without loss of generality we suppose that $\binom{w}{\nu} \in R^x$. As R^x uniquely determines the symbols below the letters y and z, we know that ν is of the shape $\sigma \cdot \tau_q \cdot \odot$, for some word σ over $\{\oplus, \ominus, \odot\}$ of length $|u_p|$.

Consider now the new word w' over $\{x, y, z\}$ defined as $w' = u_1 \cdot v_q \cdot z$. We know that w' belongs to the projection of R^y but not to the projection of R^x, because u_1 has only one occurrence of x. Let ν' be the unique word such that $\binom{w'}{\nu'} \in F$. Similarly as before, $\nu' = \sigma_1 \cdot \tau' \cdot \odot$ for some word τ' over $\{\oplus, \ominus, \odot\}$ of length $|v_q|$.

Using (3.1), we know that $h(\binom{u_1}{\sigma_1}) = h(\binom{u_p}{\sigma_p})$, and therefore $\binom{u_p}{\sigma_p} \cdot \binom{v_q}{\tau'} \cdot \binom{z}{\odot} \in F$, whose projection onto $\{x, y, z\}^+$ equals w. Since $\nu = \sigma \cdot \tau_q \cdot \odot$ and $\sigma_p \cdot \tau' \cdot \odot$ are distinct words (σ and σ_p have the same length, but σ_p contains some $\otimes$ and σ does not), the word w has at least two candidates in F, which enters in contradiction with the fact that the latter relation is functional. This concludes the proof of the claim. $\square$

By symmetry, let us assume that the first item of Claim 3.7 holds, *i.e.* for all $\binom{u_1}{\sigma_1} \in L_1^x$ and $\binom{u_p}{\sigma_p} \in L_p^x$ with $p \geq 2$, we have $h(\binom{u_1}{\sigma_1}) \neq h(\binom{u_p}{\sigma_p})$.

Claim 3.8. *The language L_1^x is in* $\mathbf{C}$.

Proof. The language $L_0^x = \left\{\binom{z}{\odot}\right\}^+$ is in $\mathbf{C}$. Therefore, $\bigsqcup_{p \geq 1} L_p^x = \{\binom{x}{\otimes}, \binom{z}{\odot}\}^+ \setminus L_0^x$ also belongs to $\mathbf{C}$. Thus, the above assumption about h-values implies the claim, because $L_1^x = h^{-1}(h(L_1^x)) \cap \bigsqcup_{p \geq 1} L_p^x$. $\square$

Now, take $\mathbb{A}_1 \subseteq \mathbb{A}_2$ as in the statement of Lemma 3.6, and consider a homomorphism g from $\mathbb{A}_2^+$ to $\left\{\binom{x}{\otimes}, \binom{z}{\odot}\right\}^+$ mapping each letter $a \in \mathbb{A}_1$ to $\binom{x}{\otimes}$ and each letter $a \in \mathbb{A}_2 \setminus \mathbb{A}_1$ to $\binom{z}{\odot}$. We have $[\exists^{=1}\mathbb{A}_1]_{\mathbb{A}_2} = g^{-1}(L_1^x)$, and, because $\mathbf{C}$ is closed under preimages under homomorphisms, $[\exists^{=1}\mathbb{A}_1]_{\mathbb{A}_2}$ is in $\mathbf{C}$. This concludes the proof of Lemma 3.6. $\square$

With a similar—yet more technical—proof, one can show that for all $p \in \mathbb{N}$, $\mathbf{C}$ also contains $[\exists^{=p}\mathbb{A}_1]_{\mathbb{A}_2}$, the language of words over $\mathbb{A}_2$ having exactly p letters in $\mathbb{A}_1$, but the result with $p = 1$ will be sufficient in the sequel of the demonstration. These results show that $\mathbf{C}$ must contain $\mathbf{FO}[\,]$, First-Order Logic with only letter tests. Recall that by Proposition 2.5 on Page 61, all relations in $\mathbf{FO}[\,]$ can be uniformised in $\mathbf{FO}[<]$. This explains why our proof of Theorem 3.1 needs to use more than once the hypothesis of $\mathbf{C}$ uniformising itself.

In the next sections, we will also use a weaker version of Lemma 3.6, when the alphabet $\mathbb{A}_1$ is composed of a unique letter:

Corollary 3.9. *Let $\mathbb{A}$ be an alphabets, and let $a \in \mathbb{A}$. Then the language $[\exists^{=1}a]_{\mathbb{A}}$, the language of finite words over $\mathbb{A}$ that contain exactly one occurrence of a, is in $\mathbf{C}$.*

3.3 Expressing subwords

Our next goal is to introduce the order $<$ on the positions of letters in a given word, meaning that $\mathbf{C}$ is capable of recognising which letters appear before the others. This is achieved gradually, with the first instance of the order expressed by the following lemma:

Lemma 3.10. *Let $\mathbb{A}$ be an alphabet and $a_0, \ldots, a_{p-1}$ be $p \geq 1$ pairwise distinct letters, that do not belong to $\mathbb{A}$. Then the language $\mathbb{A}^* \cdot a_0 \cdot \mathbb{A}^* \cdots \mathbb{A}^* \cdot a_{p-1} \cdot \mathbb{A}^*$ is in $\mathbf{C}$.*

Proof. In this proof, L denotes the language $\mathbb{A}^* \cdot a_0 \cdot \mathbb{A}^* \cdots \mathbb{A}^* \cdot a_{p-1} \cdot \mathbb{A}^*$.

We consider two distinct letters, x and y, and $p{+}1$ distinct symbols $\square, \triangle_0, \ldots, \triangle_{p-1}$, and we define the relation $R = \mathbb{C}^{\oplus}$, where $\mathbb{C}$ is the alphabet $\left\{ \left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right) \right\} \sqcup \left\{ \left(\begin{smallmatrix} x \\ \triangle_i \end{smallmatrix} \right) \mid i \in p \right\}$.

We know from Corollary 3.5 that $R \in \mathbf{C}$, and therefore it admits by assumption a uniformisation $F \in \mathbf{C}$. Let $S \in \mathbf{V}$, recognising F, let h be a homomorphism from $\mathbb{C}^+$ to S, and let $H \subseteq S$ be such that $F = h^{-1}(H)$.

We define now the word $u = y^\sharp \cdot x \cdot y^\sharp \cdots y^\sharp \cdot x \cdot y^\sharp$, where x appears exactly p times, and $\sharp$ is the natural number $\sharp(S)$. Since u is in the projection of R, it also belongs to the projection of F. Let σ be the unique word satisfying $\left(\begin{smallmatrix} u \\ \sigma \end{smallmatrix} \right) \in F$. The word σ is necessarily of the shape $\square^\sharp \cdot \triangle_{\delta(0)} \cdot \square^\sharp \cdots \square^\sharp \cdot \triangle_{\delta(p-1)} \cdot \square^\sharp$, where δ is a *permutation* of $p = \{0, \ldots, p{-}1\}$, *i.e.* a bijection from p to itself.

Let e be $h(\left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp)$. It is an idempotent element of S. Consider g the homomorphism from words over $\mathbb{A}' := \mathbb{A} \sqcup \{a_i \mid i \in p\}$ to S defined by $g(a_i) = e * h(\left(\begin{smallmatrix} x \\ \triangle_{\delta(i)} \end{smallmatrix} \right)) * e$ for $i \in p$, and $g(a) = e$ for $a \in \mathbb{A}$.

Now, consider δ' a second permutation of p, and let w be any word over $\mathbb{A}'$ of the shape $w_0 \cdot a_{\delta'(0)} \cdot w_1 \cdots w_{p-1} \cdot a_{\delta'(p-1)} \cdot w_p$, with the w_i's being arbitrary words over $\mathbb{A}$. We obtain the following equalies:

$$\begin{aligned}
g(w) &= g(w_0) \cdot g(a_{\delta'(0)}) \cdot g(w_1) \cdots g(w_{p-1}) \cdot g(a_{\delta'(p-1)}) \cdot g(w_p) \\
&= e^{|w_0|+1} \cdot h(\left(\begin{smallmatrix} x \\ \triangle_{\delta(\delta'(0))} \end{smallmatrix} \right)) \cdot e^{|w_1|+2} \cdots e^{|w_{p-1}|+2} \cdot h(\left(\begin{smallmatrix} x \\ \triangle_{\delta(\delta'(p-1))} \end{smallmatrix} \right)) \cdot e^{|w_p|+1} \\
&= e \cdot h(\left(\begin{smallmatrix} x \\ \triangle_{\delta(\delta'(0))} \end{smallmatrix} \right)) \cdot e \cdots e \cdot h(\left(\begin{smallmatrix} x \\ \triangle_{\delta(\delta'(p-1))} \end{smallmatrix} \right)) \cdot e \\
&= h(\left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp) \cdot h(\left(\begin{smallmatrix} x \\ \triangle_{\delta(\delta'(0))} \end{smallmatrix} \right)) \cdot h(\left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp) \cdots h(\left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp) \cdot h(\left(\begin{smallmatrix} x \\ \triangle_{\delta(\delta'(p-1))} \end{smallmatrix} \right)) \cdot h(\left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp) \\
&= h\Big(\left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp \cdot \left(\begin{smallmatrix} x \\ \triangle_{\delta(\delta'(0))} \end{smallmatrix} \right) \cdot \left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp \cdots \left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp \cdot \left(\begin{smallmatrix} x \\ \triangle_{\delta(\delta'(p-1))} \end{smallmatrix} \right) \cdot \left(\begin{smallmatrix} y \\ \square \end{smallmatrix} \right)^\sharp \Big)
\end{aligned}$$

Since F is functional, $g(w) \in H$ if and only if the words $\sigma = \square^\sharp \cdot \triangle_{\delta(0)} \cdot \square^\sharp \cdots \square^\sharp \cdot \triangle_{\delta(p-1)} \cdot \square^\sharp$ and $\square^\sharp \cdot \triangle_{\delta(\delta'(0))} \cdot \square^\sharp \cdots \square^\sharp \cdot \triangle_{\delta(\delta'(p-1))} \cdot \square^\sharp$ are the same words, meaning if and only if $\delta(i) = \delta(\delta'(i))$ for all $i \in p$, which is true if and only if δ' is the identity of p. Therefore,

$$g^{-1}(H) \cap \bigcap_{i \in p} [\exists^{=1} a_i]_{\mathbb{A}'} = \mathbb{A}^* \cdot a_0 \cdot \mathbb{A}^* \cdots \mathbb{A}^* \cdot a_{p-1} \cdot \mathbb{A}^* = L.$$

Using Lemma 3.9, each of the languages $[\exists^{=1} a_i]_{\mathbb{A}'}$ is in $\mathbf{C}$. Because $\mathbf{C}$ is closed under intersections, we can conclude that L is in $\mathbf{C}$. $\square$

Now we need to strengthen the above lemma, to be able to compare the positions of letters not necessarily distinct, and that may belong to the main alphabet $\mathbb{A}$:

Lemma 3.11. *Let $\mathbb{A}$ be an alphabet and let $a_0, \ldots, a_{p-1}$ be letters of $\mathbb{A}$, with $p \geq 1$. Then the language $\mathbb{A}^* \cdot a_0 \cdot \mathbb{A}^* \cdots \mathbb{A}^* \cdot a_{p-1} \cdot \mathbb{A}^*$ is in $\mathbf{C}$.*

The language in the lemma is denoted by $[\exists a_0 {<} a_1 {<} \ldots {<} a_{p-1}]_{\mathbb{A}}$.

As stated in the introduction of this chapter, this lemma is equivalent to saying that $\mathbf{C}$ contains all Boolean combinations of existential First-Order sentences with the order.

Proof. Let $\mathbb{B} := \{\triangle_0, \ldots, \triangle_{p-1}, \square\}$ be an alphabet containing $p{+}1$ pairwise distinct symbols. First, we consider the following relation:

$$R = \binom{\mathbb{A}}{\square}^* \cdot \binom{a_0}{\triangle_0} \cdot \binom{\mathbb{A}}{\square}^* \cdots \binom{\mathbb{A}}{\square}^* \cdot \binom{a_{p-1}}{\triangle_{p-1}} \cdot \binom{\mathbb{A}}{\square}^*$$

It is immediate to see that R is functional, and moreover that $\Pi_{\mathbb{A}^+}(R)$ is exactly the language $[\exists a_0 {<} \ldots {<} a_{p-1}]_{\mathbb{A}}$. Consider the relations $R_1 := R \cdot \binom{\bullet}{\triangleleft} \cdot \binom{\mathbb{A}}{\square}^*$ and $R_2 := \binom{\mathbb{A}}{\square}^* \cdot \binom{\bullet}{\triangleright} \cdot R$, where $\bullet$ is a letter not in $\mathbb{A}$, and $\triangleleft$, $\triangleright$ are two symbols both not in $\mathbb{B}$.

To conclude the proof of Lemma 3.11, we will use a fairly technical lemma. It may be seen as an abstract generalisation of the technique used in the proof of Claim 2.8 in the previous chapter:

Lemma 3.12. *Let T be a relation of finite words over a product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$. Let $\bullet$, $\square$, $\triangleleft$, and $\triangleright$ be four distinct symbols, with $\bullet$ not belonging to $\mathbb{A}$, and $\triangleleft$, $\triangleright$ both not belonging to $\mathbb{B}$. We define T_1 and T_2 as the two relations $T \cdot \binom{\bullet}{\triangleleft} \cdot \binom{\mathbb{A}}{\square}^*$ and $\binom{\mathbb{A}}{\square}^* \cdot \binom{\bullet}{\triangleright} \cdot T$ respectively, and $T_\sqcup$ as the union $T_1 \sqcup T_2$. Then, if $T_\sqcup$ is in $\mathbf{C}$ then so is $\Pi_{\mathbb{A}^+}(T)$.*

Proof. Suppose that the relation $T_\sqcup$ belongs to $\mathbf{C}$. Let $F \in \mathbf{C}$ be a uniformisation of it, recognised by $S \in \mathbf{V}$ via a homomorphism h from $\binom{\mathbb{A} \sqcup \{\bullet\}}{\mathbb{B} \cup \{\square, \triangleleft, \triangleright\}}^+$ to S, and let $H \subseteq S$ be such that $F = h^{-1}(H)$.

Let g be the homomorphism from $\mathbb{A}^+$ to S defined by $g(a) = h(\binom{a}{\square})$, for every $a \in \mathbb{A}$. We define L as $\Pi_{\mathbb{A}^+}(T)$. Notice that if for all words w_1, w_2 in $\mathbb{A}^+$, the equality $g(w_1) = g(w_2)$ implies the equivalence $w_1 \in L$ iff $w_2 \in L$, then $L = g^{-1}\big(g(L)\big)$, and therefore $L \in \mathbf{C}$.

We show that this implication holds indeed for all $w_1, w_2 \in \mathbb{A}^+$. Suppose that there exist two words $w_1 \in L$ and $w_2 \in L^{\mathrm{c}}$ such that $g(w_1) = g(w_2)$, in order to provide a contradiction.

Let w be the word $w_1 \cdot \bullet \cdot w_1$, over the alphabet $\mathbb{A} \sqcup \{\bullet\}$. This word w is in both $\Pi_{(\mathbb{A}\sqcup\{\bullet\})^+}(T_1)$ and $\Pi_{(\mathbb{A}\sqcup\{\bullet\})^+}(T_2)$, and therefore in $\Pi_{(\mathbb{A}\sqcup\{\bullet\})^+}(T_\sqcup)$. Let σ be the unique word over $\mathbb{B} \cup \{\square, \triangleleft, \triangleright\}$ such that $\binom{w}{\sigma} \in F$. This pair of words is either in T_1 or in T_2.

We suppose for instance that $\binom{w}{\sigma} \in T_1$ (the case $\binom{w}{\sigma} \in T_2$ is symmetric). Because $\bullet$ is not in $\mathbb{A}$, σ is necessarily of the shape $\sigma_1 \cdot \triangleleft \cdot \square^{|w_1|}$, with $\sigma_1 \in \mathbb{B}^{|w_1|}$ such that $\binom{w_1}{\sigma_1} \in T$.

Let w' be the word $w_2 \cdot \bullet \cdot w_1$, also over $\mathbb{A} \sqcup \{\bullet\}$, and also in $\Pi_{(\mathbb{A}\sqcup\{\bullet\})^+}(T_2) \subseteq \Pi_{(\mathbb{A}\sqcup\{\bullet\})^+}(T_\sqcup)$. Let σ' be the unique word over $\mathbb{B} \cup \{\square, \triangleleft, \triangleright\}$ such that $\binom{w'}{\sigma'} \in F$. Again, σ' is necessarily of the shape $\square^{|w_2|} \cdot \triangleright \cdot \sigma'_1$, with $\sigma'_1 \in \mathbb{B}^{|w_1|}$ such that $\binom{w_1}{\sigma'_1} \in T$. Since $g(w_2) = g(w_1)$, we know that $h\Big(\binom{w_2}{\square^{|w_2|}} \cdot \binom{\bullet}{\triangleright} \cdot \binom{w_1}{\sigma'_1}\Big) = h\Big(\binom{w_1}{\square^{|w_1|}} \cdot \binom{\bullet}{\triangleright} \cdot \binom{w_1}{\sigma'_1}\Big)$. The latter value does not belong to H because $\binom{w_1}{\square^{|w_1|}} \cdot \binom{\bullet}{\triangleright} \cdot \binom{w_1}{\sigma'_1} \notin F$—we know that F is functional and $\sigma \neq \square^{|w_1|} \cdot \triangleright \cdot \sigma'_1$. This means that $\binom{w'}{\sigma'}$ is not in F, contradicting the assumption, and concluding the proof of this lemma. $\square$

Now we go back to the proof of Lemma 3.11. The letters $\binom{a_0}{\triangle_0}, \ldots, \binom{a_{p-1}}{\triangle_{p-1}}, \binom{\bullet}{\triangleleft}, \binom{\bullet}{\triangleright}$ are all pairwise distinct, and none of them is in the alphabet $\binom{\mathbb{A}}{\{\square\}}$. Therefore, Lemma 3.10 tells us that R_1 and R_2 are in $\mathbf{C}$. This means that the union relation $R_\sqcup = R_1 \sqcup R_2$ is in $\mathbf{C}$, and we can conclude with Lemma 3.12 that $[\exists a_0 < \ldots < a_{p-1}]_{\mathbb{A}} = \Pi_{\mathbb{A}^+}(R)$ is in $\mathbf{C}$. $\square$

Corollary 3.13. *Let $\mathbb{A}_0, \ldots, \mathbb{A}_{p-1}$ be pairwise disjoint alphabets, with p being a positive natural number. Then the language $\mathbb{A}_0^* \cdot \mathbb{A}_1^* \cdots \mathbb{A}_{p-1}^* \setminus \{\epsilon\}$ is in $\mathbf{C}$.*

Proof. It is enough to observe that this language is exactly $\bigcap_{i \in j \in p} \bigcap_{a_i \in \mathbb{A}_i} \bigcap_{a_j \in \mathbb{A}_j} [\exists a_j < a_i]^{\mathrm{c}}_{\mathbb{A}}$ (where $\mathbb{A}$ is the union of the $\mathbb{A}_i$'s). $\square$

3.4 Evaluating words in semigroups

We will now prove a variant of Lemma 3.11 for *polynomials*. A *monomial* over an alphabet $\mathbb{A}$ is a language of the shape $L_0 \cdot L_1 \cdots L_{p-1}$, where each L_i is either of the shape $\mathbb{A}_i^*$, with $\mathbb{A}_i$ being a subset of $\mathbb{A}$, or a language having a single word over $\mathbb{A}$, of length one, and such that at least one of the L_i's is of the latter kind (in order to avoid the empty word). Notice that, in the former kind, the subset $\mathbb{A}_i$ may be empty, and in this case, $\mathbb{A}_i^* = \{\epsilon\}$, as stated on page 25. An example of a monomial is the language $\{a,b\}^* \cdot \{d\} \cdot \{c\}^*$, which we simply write $\{a,b\}^* \cdot d \cdot \{c\}^*$.

For convenience, we can write that a monomial is of the shape $\mathbb{A}_0^{\xi_0} \cdot \mathbb{A}_1^{\xi_1} \cdots \mathbb{A}_1^{\xi_{p-1}}$, with each ξ being either the symbol $*$, or 1, and $\mathbb{A}_i$ being a singleton in the latter case. Notice that the alphabets $\mathbb{A}_i$ are not required to be pairwise disjoint in that definition.

We define a *polynomial* as being a finite union of monomials.

Remark 3.14. *The family of polynomials is closed under unions and concatenations.*

Proof. The closure under unions follows directly from definition. The closure under concatenations follows from the fact that the concatenation is *distributive* over the union, meaning that for any languages L_1, L_2, and L, we have $L \cdot (L_1 \cup L_2) = (L \cdot L_1) \cup (L \cdot L_2)$ and $(L_1 \cup L_2) \cdot L = (L_1 \cdot L) \cup (L_2 \cdot L)$. $\square$

The results of the previous sections give us the ingredients to prove that polynomials are elements of $\mathbf{C}$.

Lemma 3.15. *Any polynomial is in* $\mathbf{C}$.

Proof. First, since $\mathbf{C}$ is closed under unions, it is enough to prove the lemma for monomials. Consider a monomial L over an alphabet $\mathbb{A}$, *i.e.* $L = \mathbb{A}_0^{\xi_0} \cdot \mathbb{A}_1^{\xi_1} \cdots \mathbb{A}_{p-1}^{\xi_{p-1}}$, where each ξ_i is understood either as $*$ or as 1, with $\mathbb{A}_i$ being a singleton $\{a_i\}$ in the latter case. We define $\mathbb{A}'$ as the alphabet $\mathbb{A} \sqcup \{\bullet\}$, with $\bullet$ not being in $\mathbb{A}$, and $\mathbb{B}$ as the alphabet $\{\Delta_0, \ldots, \Delta_{p-1}, \square, \triangleleft, \triangleright\}$, all these symbols being distinct. Let R be the relation $\binom{\mathbb{A}_0}{\Delta_0}^{\xi_0} \cdots \binom{\mathbb{A}_{p-1}}{\Delta_{p-1}}^{\xi_{p-1}}$, whose projection onto $\mathbb{A}^+$ is L, and let R_1, R_2 be the relations $R \cdot \binom{\bullet}{\triangleleft} \cdot \binom{\mathbb{A}}{\square}^*$ and $\binom{\mathbb{A}}{\square}^* \cdot \binom{\bullet}{\triangleright} \cdot R$ respectively.

Notice that:

$$R_1 = \binom{\mathbb{A}_0}{\Delta_0}^* \cdots \binom{\mathbb{A}_{p-1}}{\Delta_{p-1}}^* \cdot \binom{\bullet}{\triangleleft}^* \cdot \binom{\mathbb{A}}{\square}^* \;\cap\; \left[\exists^{=1}\binom{\bullet}{\triangleleft}\right]_{\binom{\mathbb{A}'}{\mathbb{B}}} \;\cap\; \bigcap_{i \in p} T_i, \tag{3.3}$$

where, for each $i \in p$, the language T_i is either: $\binom{\mathbb{A}'}{\mathbb{B}}^+$ if ξ_i is $*$; or $\left[\exists^{=1}\binom{a_i}{\Delta_i}\right]_{\binom{\mathbb{A}'}{\mathbb{B}}}$ if ξ_i is 1. Now, the first ingredient on the right-hand side of (3.3) is as in Corollary 3.13 and thus belongs to $\mathbf{C}$. The second ingredient also is in $\mathbf{C}$, by Corollary 3.9, and so is each T_i: by Lemma 3.3 if it is $\binom{\mathbb{A}'}{\mathbb{B}}^+$, and by Corollary 3.9 again if it is $\left[\exists^{=1}\binom{a_i}{\Delta_i}\right]_{\binom{\mathbb{A}'}{\mathbb{B}}}$. Therefore, R_1 is in $\mathbf{C}$, and similarly, we have $R_2 \in \mathbf{C}$. Thus, Lemma 3.12 implies that $L = \Pi_{\mathbb{A}^+}(R) \in \mathbf{C}$, which concludes the proof of Lemma 3.15. $\square$

We can now conclude the proof of Theorem 3.1. Let L be a regular language over some alphabet $\mathbb{A}$, recognised by a finite semigroup S (*a priori* not in $\mathbf{V}$) via the automorphism h from $\mathbb{A}^+$ to S: $L = h^{-1}(H)$ with some $H \subseteq S$. We dedicate the end of the section to the proof that $L \in \mathbf{C}$.

Consider a pair of words $\binom{w}{\gamma} \in \binom{\mathbb{A}}{S}^+$ of length n. We say that such a word is an *evaluation* if for every $i \in n$ we have $\gamma(i) = h\big(w(0)\cdots w(i)\big)$. Notice that, in this case, $w \in L$ if and only if $\gamma(n{-}1) \in H$. Let E be the set of words $\binom{w}{\gamma} \in \binom{\mathbb{A}}{S}^+$ that are evaluations and such that $\gamma(n{-}1) \in H$.

Claim 3.16. *We have* $\Pi_{\mathbb{A}^+}(E) = L$.

Proof. For each word $w \in \mathbb{A}^n$ with $n \geq 1$, there exists a unique word $\gamma \in S^n$ such that $\binom{w}{\gamma}$ is an evaluation. In that case, $h(w) = \gamma(n{-}1)$. Thus, $w \in L$ iff $\gamma(n{-}1) \in H$ iff $\binom{w}{\gamma} \in E$. $\square$

Our aim is to show that a variant of the set of evaluations E belongs to $\mathbf{C}$ and then invoke Lemma 3.12 to conclude that L is in $\mathbf{C}$.

Consider $a, b \in \mathbb{A}$ and $r, s \in S$ and define

$$I_{a,r} := \binom{a}{r} \cdot \binom{\mathbb{A}}{S}^*, \quad M_{a,r,b,s} := \binom{\mathbb{A}}{S}^* \cdot \binom{a}{r} \cdot \binom{b}{s} \cdot \binom{\mathbb{A}}{S}^*, \quad F_{a,r} := \binom{\mathbb{A}}{S}^* \cdot \binom{a}{r}.$$

Let W be the union of the following languages: the languages $I_{a,r}$ ranging over $a \in \mathbb{A}$ and $r \in S$ such that $h(a) \neq r$; the languages $M_{a,r,b,s}$ ranging over $a, b \in \mathbb{A}$ and $r, s \in S$ such that $r * h(b) \neq s$; and the languages $F_{a,r}$ ranging over $r \notin H$. Notice that W, as a union of polynomials, is a polynomial.

Claim 3.17. *The complement of W is exactly E.*

Proof. Clearly $E \cap W = \emptyset$. Thus, it is enough to prove that if $\binom{w}{\gamma} \notin W$ then $\binom{w}{\gamma} \in E$. Let $n = |w| = |\gamma|$. Since $\binom{w}{\gamma} \notin W$, we know that $\gamma(n{-}1) \in H$ (see the languages $F_{a,r}$), thus it is enough to show that $\binom{w}{\gamma}$ is an evaluation. We prove it inductively over $i \in n$. The fact that $\gamma(0) = h\big(w(0)\big)$ follows from the assumption that $\binom{w}{\gamma} \notin W$ (see the languages $I_{a,r}$).

Take $i < n{-}1$ and assume that $\gamma(i) = h\big(w(0)\cdots w(i)\big)$. Observe that $\gamma(i{+}1)$ must be equal to $\gamma(i) * h\big(w(i{+}1)\big)$ (see the languages $M_{a,r,b,s}$). Thus, $\gamma(i{+}1) = h\big(w(0)\cdots w(i{+}1)\big)$. This concludes the proof. $\square$

Consider distinct letters $\bullet$, $\square$, $\triangleleft$, $\triangleright$, all neither in $\mathbb{A}$ nor in S, and let $\mathbb{A}'$ and S' be the alphabets $\mathbb{A} \sqcup \{\bullet\}$ and $S \sqcup \{\square, \triangleleft, \triangleright\}$ respectively. We define the following four languages of finite words over $\binom{\mathbb{A}'}{S'}$:

$$
\begin{aligned}
R_1 &:= W \cdot \tbinom{\bullet}{\triangleleft} \cdot \tbinom{\mathbb{A}}{\square}^*, & R_2 &:= \tbinom{\mathbb{A}}{\square}^* \cdot \tbinom{\bullet}{\triangleright} \cdot W,\\
R_1' &:= R_1^{\mathrm{c}} \ \cap \ \tbinom{\mathbb{A}}{S}^* \cdot \tbinom{\bullet}{\triangleleft} \cdot \tbinom{\mathbb{A}}{\square}^*, & R_2' &:= R_2^{\mathrm{c}} \ \cap \ \tbinom{\mathbb{A}}{\square}^* \cdot \tbinom{\bullet}{\triangleright} \cdot \tbinom{\mathbb{A}}{S}^*.
\end{aligned}
$$

Notice that both R_1 and R_2 are polynomials (see Remark 3.14).

Claim 3.18. *Using the above notions, we have* $R_1' = E \cdot \binom{\bullet}{\triangleleft} \cdot \binom{\mathbb{A}}{\square}^*$ *and* $R_2' = \binom{\mathbb{A}}{\square}^* \cdot \binom{\bullet}{\triangleright} \cdot E$.

Proof. These equalities follow directly from definition and Claim 3.17. $\square$

By Lemma 3.15, the languages R_1 and R_2 are in $\mathbf{C}$, since they are polynomials. Now, because $\mathbf{C}$ is closed under Boolean operations, R_1', R_2', and their union $R_\sqcup = R_1' \sqcup R_2$ are also in $\mathbf{C}$. Therefore, Lemma 3.12 guarantees that $\Pi_{\mathbb{A}^+}(E) \in \mathbf{C}$. Thus, by Claim 3.16 we know that $L \in \mathbf{C}$. We can finally conclude Theorem 3.1:

Theorem 3.1. $\mathbf{MSO}[<]$ *is the unique non-empty variety of languages that satisfies the uniformisation property.*

3.5 Conclusion and further questions

In this chapter, we were able to prove that, among all the non-empty varieties of languages, $\mathbf{MSO}[<]$, the class of all regular languages, can be characterised by the property of uniformising its own relations. In other words, the second-order quantifiers $\exists X$ are a crucial tool in order to construct uniformisations. This is another argument for the very natural character of $\mathbf{MSO}[<]$.

Nevertheless, we could wonder if this result holds when considering less restrictive closure properties than the one of varieties of languages: for example, what about the classes that are also closed under Boolean combinations, under preimages under homomorphisms, but not necessarily under quotients?

Indeed, when relying on Eilenberg's correspondence between varieties of languages and varieties of semigroups, we implicitly make use of the closures under quotients. Yet, along our proof, we never namely use this assumption, in contrast to the closures under Boolean combinations and preimages under homomorphisms. Therefore, we could wonder if it is a less important assumption, and if $\mathbf{MSO}[<]$ would also be the unique non-empty class of languages closed under Boolean combinations, preimages of homomorphisms, and satisfying the uniformisation property.

Question 3.19. *Is* **MSO**$[<]$ *the unique non-empty self-uniformising class of languages which is closed under Boolean combinations and preimages under homomorphisms?*

Another interesting question is whether our theorem remains true if we consider closures under preimages under not all homomorphisms, but only a certain class of them. For instance, there exists a field of research studying classes closed under preimages of *length-preserving* homomorphisms: those are the homomorphisms induced from a function $\mathbb{A} \to \mathbb{B}$. One also meets another notion of *length-multiplying* homomorphisms: those are the homomorphisms induced from a function $\mathbb{A} \to \mathbb{B}^k$, where k is some positive natural number. We refer to [Pin12] for a more complete descriptions of these homomorphisms and of the classes of languages closed by their preimages.

Question 3.20. *Is* **MSO**$[<]$ *the unique non-empty self-uniformising class of regular languages which is closed under Boolean combinations, left and right quotients, and preimages under length-preserving homomorphisms?*

Question 3.21. *Is* **MSO**$[<]$ *the unique non-empty self-uniformising class of regular languages which is closed under Boolean combinations, left and right quotients, and preimages under length-multiplying homomorphisms?*

Our proof seems to crucially use the assumption that our variety $\mathbf{C}$ of languages is closed under preimages under any homomorphism. Indeed, in our proofs of Lemmata 3.2 and 3.10, the $\sharp$-numbers of semigroups are involved, and these are *a priori* not bounded. That would suggest a negative answer to Questions 3.20 and 3.21. However, the author and his supervisors were not able to construct an example of a second class satisfying all these properties.

More generally, Questions 3.19 to 3.21 are particular instances of the natural problem:

Problem 3.22. *Considering some closures of classes of languages, weaker than the closures of varieties of languages. Is* **MSO**$[<]$ *the unique class of languages satisfying these closures and uniformising its own relations?*

Figure 3.3 is an illustration of Problem 3.22: the dashed ellipse depicts the classes of languages satisfying these weaker closures, and we wonder if it does intersect with the green ellipse in some other class than **MSO**$[<]$.

Since we know that **MSO**$[<]$ uniformises itself in the case of ω-words (see Theorem 1.49 on page 58), another natural question arising from Theorem 3.1 is whether it still holds when considering said ω-words. In this field of research, the natural robustness to consider

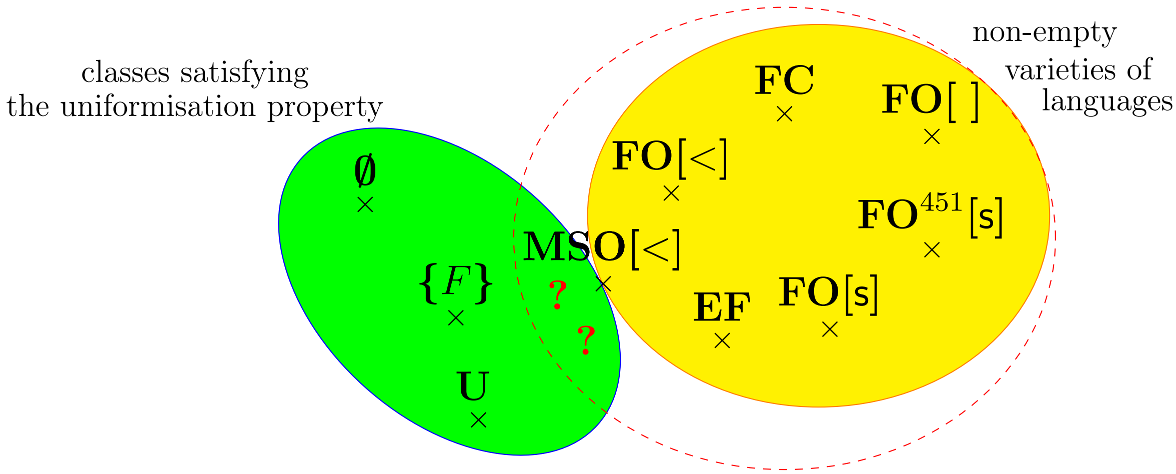


Figure 3.3: Do these ellipses intersect in some other class than **MSO**[<]?

is the robustness of ω-*varieties of languages*, or ∞-*varieties of languages*. Algebraically, they correspond to varieties of finite *Wilke algebrae*. A survey of these notions, which we do not develop in the thesis, can be found in [CPP08, Section 3.3].

Question 3.23. *Is* **MSO**[<] *the unique non-empty* ω*-variety of languages that uniformise its own relations?*

At the moment of writing these lines, the author was not able to answer this question, neither positively nor negatively. If **MSO**[<] is indeed the unique such non-empty ω-variety, the strategy to prove it shall follow a pattern similar to this chapter: consider any non-empty such ω-variety, and use both its uniformisation property and its closures to show that it can express more and more properties of ω-words, until we ultimately show that it necessarily contains all the regular ω-languages.

During this process, the author was able to adapt the results of Section 3.1, and to show that any non-empty self-uniformising ω-variety of languages necessarily has the capacity of testing if an ω-word contains certain letters, and also of testing if it contains these letters a certain number of times. Yet, it seems that, in order to go further, we need to prove also that our ω-variety is moreover capable of testing if a letter appears in an ω-word infinitely many times. The author did not succeed in this seemingly crucial step yet.

Chapter 4

Regular uniformisations on finitary linear orders

In this chapter, we give a characterisation of finitary linear orders on which regular relations admit uniformisations that are also regular. Recall that a finitary linear order is a countable linear order that is obtained from singleton sets using a finite number of times the operations $+$, $\times\omega$, $\times\omega^\star$, and η (see page 27 for a complete definition of these operations), and therefore admits a finite representation. Furthermore, finitary linear orders and words are convenient since it is decidable whether they satisfy a given MSO[$<$] formula. This is a direct corollary of Theorem 26 in [CCP18], and also of Theorem 6.2 of [She75].

We characterise the possibility of uniformising regular relations by the non-existence of non-trivial automorphisms. In this sense, the theorem, originally proved in [MS20], can be seen as a generalisation of Theorem 1.49 on page 58, that states that the linear order ω satisfies the regular-uniformisation property. In fact, a part of the proof we propose relies on this theorem. In addition, we give other equivalent conditions to this regular-uniformisation property, such as the existence of regular choice functions.

In Section 4.1, we state the main theorem of this chapter. We additionally explain how a stronger version can be obtained, where relations are defined not over a given finitary linear order, but over a given finitary word. In Section 4.2, we prove the most direct implications of our theorem. In Section 4.3, we display a certain number of objects initially introduced in [CCP18], such as condensation trees and tree decompositions, and we explain how they relate to automorphisms of finitary linear orders. In Section 4.4, we finally prove that a finitary linear order satisfies the regular-uniformisation property when it is without non-trivial automorphisms. In the last section, we show how to express, in MSO, and under the same assumption, related objects, such as regular choice functions.

4.1 Stating the theorem

In this section, we give a full formulation of the main theorem of the chapter, and explicit some notions and notations involved in the formulation. This section is divided into two subsection, each one stating a version of the theorem: the first one being about finitary linear orders, the second one about finitary words.

4.1.1 Main statement

Here is the statement of the main theorem of this chapter:

Theorem 4.1. *Let λ be a finitary linear order. The following conditions are equivalent:*

i) *λ is* rigid *(i.e. it does not admit any non-trivial automorphism),*

ii) *λ does not have any convex subset isomorphic to $\mu \times \mathbb{Z}$, for any non-empty linear order μ,*

iii) *λ satisfies the regular-uniformisation property,*

iv) *λ admits a regular choice function,*

v) *λ admits a regular well order,*

vi) *each position of λ can be* regularly defined*.*

Moreover, Items iii) to vi) are effective: there is a procedure that inputs a representation of a regular relation $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^{\lambda}$, and outputs a representation of a regular uniformisation of it if there exists one, and **NO** *if the above conditions fail for λ. And there is a similar procedure for choice functions, for well orders, and for definitions of the positions.*

Let us define precisely the notions involved in the theorem.

As defined on page 23 in Chapter 1, an automorphism of λ is a bijective function α from λ to itself such that for all $x, y \in \lambda$, $\alpha(x) < \alpha(y)$ if and only if $x < y$. It is called non-trivial if it is not the identity function $\mathrm{id}_\lambda\colon x \mapsto x$. For instance, the function $x \mapsto x{+}1$ is a non-trivial automorphism of the linear order $\mathbb{Z}$.

We say that λ is *rigid* if id_λ is its unique automorphism. Hence, linear orders that admit non-trivial automorphisms, like $\mathbb{Z}$, are called *non-rigid.*

We can give ω and $\omega^\star$ as examples of rigid linear orders:

Example 4.2. *The linear order ω is rigid.*

Proof. Let α be an automorphism of ω. We define X as the subset $\{n \in \mathbb{N} \mid \alpha(n) \neq n\}$, and show that it is necessarily the empty set.

In order to reach a contradiction, let us suppose that X is not empty. Since ω is a well order, X must have some least element m, which, by definition, is such that $\alpha(m) \neq m$.

If $\alpha(m) < m$, then, $\alpha(m)$ is not in X (because m is its least element), and therefore α maps it to itself: $\alpha(\alpha(m)) = \alpha(m)$. Hence, $\alpha(m)$ and m are two distinct positions of ω, mapped to the same element, which contradicts the assumption that α is injective. Therefore, we necessarily have $m < \alpha(m)$.

But now, we realise that α does not map any element to m. Indeed, any $n < m$ is mapped to itself by assumption, and if $n \geq m$, then we have $\alpha(n) \geq \alpha(m) > m$ because α preserves the order. Therefore, the very existence of m enters into contradiction with the surjectivity of α, and we have reached our point: X is empty, and α is the identity. $\square$

In Item ii), $\mu \times \mathbb{Z}$ has to be understood as a simplified notation for $\mu \times \omega^\star + \mu \times \omega$: it consists of a bi-infinite number of copies of μ. For instance, if μ is a singleton, or even any finite linear order, then $\mu \times \mathbb{Z}$ is isomorphic to $\mathbb{Z}$ itself.

Item iii) tells us that for all alphabets $\mathbb{A}$ and $\mathbb{B}$, for any relation $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ defined by an MSO$[\binom{\mathbb{A}}{\mathbb{B}}, <]$ formula, there exists a relation $F \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ that uniformises R and is also defined by an MSO$[\binom{\mathbb{A}}{\mathbb{B}}, <]$ formula. The linear order ω can be given as an example of a linear order satisfying this item, see Theorem 1.49 on page 58.

We defined on page 53, also in Chapter 1, that a choice function f over λ is regular if it is defined by an MSO[<] formula $\varphi^\lambda_{\text{choice}}(X, x)$ in the following sense: for any non-empty subset X of λ, there exists a unique $x \in \lambda$ such that $\lambda \vDash \varphi^\lambda_{\text{choice}}(X, x)$, and moreover said x is $f(X)$ (and therefore belongs to X). Example 1.38 in the same chapter, highlights that the hypothesis of λ being well ordered naturally induces such a formula.

In Item v), a *well order* $\preceq_{\text{wo}}$ over λ is *regular* if it is defined by an MSO[<] formula $\varphi^\lambda_{\text{wo}}(x, y)$: $\preceq_{\text{wo}}$ is the relation $\{\langle x, y\rangle \mid \lambda \vDash \varphi^\lambda_{\text{wo}}(x, y)\}$. In the case when λ is a well order itself, it suffices to take $\preceq_{\text{wo}}$ as $\leq$ (defined by the formula $\varphi^\lambda_{\text{wo}}(x, y) := x{<}y \vee x{=}y$). As an instance of a linear order that is not a well order and that however admits a regular well order, we can consider $\omega^\star$: in this case, the order $\preceq_{\text{wo}}$, defined as $x \preceq_{\text{wo}} y$ if $y \leq x$, is a regular well order.

We choose to consider said well order as non-strict, meaning that it is reflexive. This choice is not problematic since we can always consider the formula $\varphi^\lambda_{\text{wo}}(x, y) \wedge x{\neq}y$ to obtain the induced strict order $\prec_{\text{wo}}$, and, reciprocally, if $\varphi^\lambda_{\text{wo}}(x, y)$ defines any strict well order $\prec_{\text{wo}}$, then $\varphi^\lambda_{\text{wo}}(x, y) \vee x{=}y$ defines the related non-strict well order $\preceq_{\text{wo}}$.

Finally, we say that an element of x of λ is *regularly definable* if there exists an MSO[<] formula $\varphi^{\lambda,x}_{\text{def}}(y)$ such that for all $y \in \lambda$, we have $\lambda \vDash \varphi^{\lambda,x}_{\text{def}}(y)$ if and only if $x = y$. In this case we say that $\varphi_{\text{def}}^{\lambda,x}(y)$ *defines* x. Item vi) says that we can construct such a formula $\varphi_{\text{def}}^{\lambda,x}$ for each individual $x \in \lambda$. Once again, ω can be given as an example of a linear order for which this item is true:

Example 4.3. *Each position of* ω *is regularly definable.*

Proof. A natural number n can be defined as the only element of ω having exactly n *predecessors*, *i.e.* elements $m \in \omega$ such that $m < n$. The property of having at least n predecessors can be defined by the formula $\varphi_{\geq n}(x) := \exists x_0, \ldots, x_{n-1}.\ x_0 < x_1 < \cdots < x_{n-1} < x$ (if $n = 0$ then this formula is $\top$ by default), and the property of having at most n predecessors is defined by the formula $\neg\varphi_{\geq n+1}(x)$. Thus, the formula $\varphi_{\text{def}}^{\omega,n}(x)$ obtained as $\varphi_{\geq n}(x) \wedge \neg\varphi_{\geq n+1}(x)$, defines n. $\square$

Theorem 4.1 makes a link between all these notions. In the original article [MS20], there is no mention of Items v) and vi), as their equivalence with the first four items was discovered later. In fact, the author thanks and pays credit to the unknown reviewer who suggested in their comments that the construction involved in the paper could be adjusted to also deduce a construction of a regular well order.

4.1.2 A stronger version involving labellings

Before proving the theorem, we state a stronger version of it, in which the involved linear order λ comes with a labelling, *i.e.* when considering a finitary word w (see page 1.1.3 for the definition of them):

Theorem 4.4. *Let* w *be a finitary word. The following conditions are equivalent:*

i) w *is rigid,*

ii) w *does not have any factor isomorphic to* $u^{\mathbb{Z}}$*, for some non-empty word* u*,*

iii) w *satisfies the regular-uniformisation property,*

iv) w *admits a regular choice function,*

v) w *admits a regular well order,*

vi) each position of w *is regularly definable.*

We shall make explicit these readjusted notions. Let $\mathbb{A}$ be the finite alphabet of w.

Item i) of Theorem 4.4 is weaker than the related item in Theorem 4.1 because it involves automorphisms not only of linear orders but of words, *i.e.* functions that preserve not only the order but also the labels (see the definition on page 24). As for linear orders, we say that w is *rigid* if the identity is its unique automorphism. It is possible for a finitary word

w to be rigid, while $\mathrm{Dom}(w)$ on the contrary *does* admit some non-trivial automorphism (of linear order).

Example 4.5. *Let $\mathbb{A}$ be an alphabet with at least two different letters: a and b. Then the finitary word w over the alphabet $\mathbb{A}$, of domain $\mathbb{Z}$, and obtained as $a^{\omega^\star} \cdot b \cdot a^\omega$, is rigid.*

Proof. Let α be an automorphism of w, and let x_0 be the unique position of $\mathbb{Z}$ labelled by b in w.

Since α preserves the labels, $\alpha(x_0)$ is also labelled by b in w, which implies that it is necessarily x_0 itself.

Then, α conserving the order, we have, for every $x \in \mathbb{Z}$, the equivalences $x_0 < x$ iff $\alpha(x_0) < \alpha(x)$ iff $x_0 < \alpha(x)$. This means that $\alpha\restriction_{>x_0}$, the function obtained by restricting α to the positions of $\mathbb{Z}$ greater than x_0, is an automorphism of $w\restriction_{>x_0}$, the word obtained by restricting w to these positions. In particular, $\alpha\restriction_{>x_0}$ is an automorphism (of linear order) of $\mathrm{Dom}(w\restriction_{>x_0})$ that is isomorphic to ω. Hence, Example 4.2 tells us that this isomorphism $\alpha\restriction_{>x_0}$ is necessarily the identity of $\mathrm{Dom}(w\restriction_{>x_0})$: for all $x > x_0$, we have $\alpha(x) = x$.

The equality $\alpha(x) = x$ for all $x < x_0$ is proven symmetrically, and this concludes the proof that α is necessarily the identity function. $\square$

We recall the definition of a factor of w: it is the word w restricted to a convex subset of its domain (see page 24). Considering that, Item ii) is exactly the same as its related item in Theorem ii), modulo the labels: $u^{\mathbb{Z}}$ is a simplified notation for $u^{\omega^\star} \cdot u^\omega$.

Items iv) to vi) are similar to their corresponding items in Theorem 4.1, with the difference that the considered formulae are now in the class $\mathrm{MSO}[\mathbb{A}, <]$: they have access to the labels of w. As an example, we show that the same word $a^{\omega^\star} \cdot b \cdot a^\omega$ satisfies Item v):

Example 4.6. *Let $\mathbb{A}$ be an alphabet with at least two different letters: a and b. Then the finitary word w over the alphabet $\mathbb{A}$, of domain $\mathbb{Z}$, and obtained as $a^{\omega^\star} \cdot b \cdot a^\omega$, admits a regular well order.*

Proof. Let x_0 be the unique position of $\mathbb{Z}$ labelled by b in $\mathbb{Z}$.

We consider the linear order $\preceq_{\text{wo}}$ on λ defined by $x < y$ if one of the three conditions is true: $x_0 \leq x < y$, $y < x_0 \leq x$, or $y < x < x_0$. The fact that $\preceq_{\text{wo}}$ is a well order results from the facts that ω is a well order, and that $\omega^\star$ with its *reversed order* $\leq^\star$ ($x \leq^\star y$ if $y \leq x$) is also one. Now, we can define x_0 in $\mathrm{MSO}[\mathbb{A}, <]$, as it is the unique position labelled by b, and therefore, the three conditions above can be defined in the same class: $\preceq_{\text{wo}}$ is regular. $\square$

Item iii) is more technical. Let $\mathbb{B}$ and $\mathbb{C}$ be two alphabets. We say that a relation $R \subseteq \binom{\mathbb{B}}{\mathbb{C}}^{\mathrm{Dom}(w)}$ is *regular over* w if there exists some $\mathrm{MSO}[\begin{pmatrix}\mathbb{A}\\\mathbb{B}\\\mathbb{C}\end{pmatrix}, <]$ fomula φ such that $R = \left\{\binom{\sigma}{\kappa} \in \binom{\mathbb{B}}{\mathbb{C}}^{\mathrm{Dom}(w)} \mid \begin{pmatrix}w\\\sigma\\\kappa\end{pmatrix} \vDash \varphi\right\}$. Notice that it does not imply *a priori* that it is regular in the common sense. Item iii) tells us that for any $\mathrm{MSO}[\begin{pmatrix}\mathbb{A}\\\mathbb{B}\\\mathbb{C}\end{pmatrix}, <]$ formula, the induced relation R is uniformised by some relation $F \subseteq \binom{\mathbb{B}}{\mathbb{C}}^{\mathrm{Dom}(w)}$ that can be also defined by some $\mathrm{MSO}[\begin{pmatrix}\mathbb{A}\\\mathbb{B}\\\mathbb{C}\end{pmatrix}, <]$ formula.

4.1.3 Overview of the proof

The next sections of the chapter are dedicated to the proof of Theorem 4.1. A proof of Theorem 4.4 is obtained from it via minor changes, *i.e.* by considering some additional labellings in the formulae we construct.

In Section 4.2, we prove the most straightforward implications of the theorem, which can be considered as folklore: the equivalence of Items *i*) and *ii*), as well as the implication of Item *i*) by Items *iii*) to *vi*).

In Section 4.4, we show how *evaluation trees*, a notion which we define in Section 4.3, can be used to construct, under the assumption that the finitary linear order λ is rigid, regular uniformisations for regular relations over λ.

Finally, in Section 4.5, we provide an algorithm to construct, again under the same assumption, a regular choice function over λ, a regular well order, and $\mathrm{MSO}[<]$ formulae defining positions of λ.

4.2 Proving the first implications of the theorem

In this section, we consider a linear order λ and prove the more direct implications of Theorem 4.1, such as the impossibility to define some properties in Monadic Second-Order Logic when λ is non-rigid. Figure 4.1 draws a complete picture of these implications. We will notice that the assumption of λ being finitary is not needed in any of these proofs.

First, we show the equivalence between non-rigidity and convex subsets isomorphic to the linear order $\mu \times \mathbb{Z}$, for some non-empty linear order μ:

Claim 4.7. *The following two conditions are equivalent:*

$\neg i$) λ is non-rigid,

$\neg ii$) some convex subset of λ is isomorphic to $\mu \times \mathbb{Z}$, with μ being some non-empty linear order.

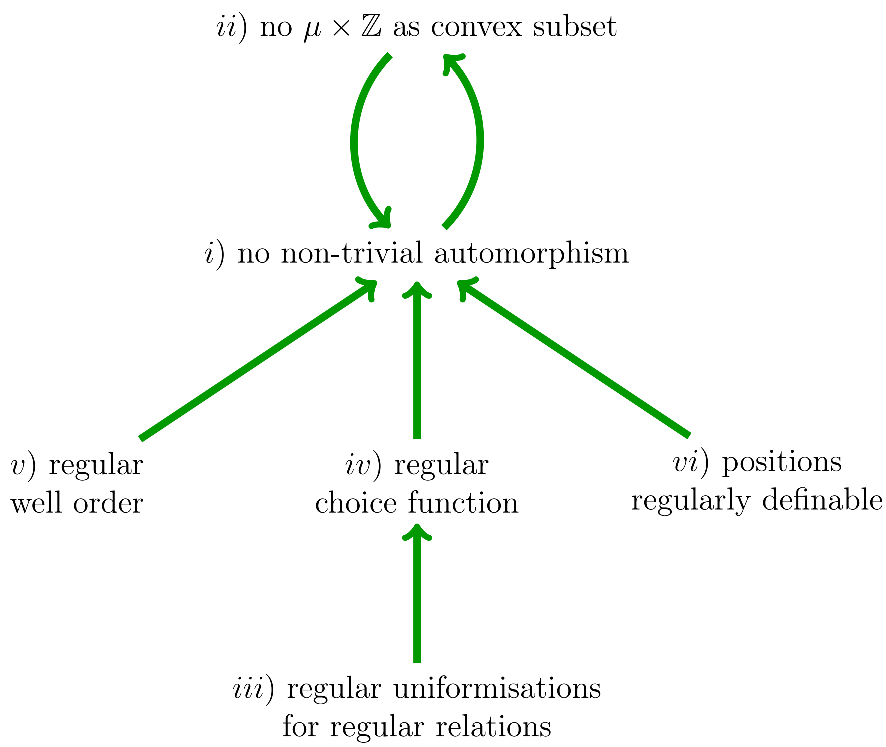


Figure 4.1: None of these implications require the linear order to be finitary.

Proof. First, let us suppose that λ admits a non-trivial automorphism α. For $k \in \mathbb{N}$ and $x \in \lambda$, we naturally define $\alpha^k(x)$ as the element $\alpha(\alpha(\ldots\alpha(x)))$, where the function α is applied exactly k times to x (if $k = 0$ then the resulting element is x by convention), and $\alpha^{-k}(x)$ as the element $(\alpha^{-1})^k(x)$.

Let $x_0 \in \lambda$ be a position such that $\alpha(x_0) \neq x_0$ (there must exist one, by the assumption). Without loss of generality, we can suppose that $x_0 < \alpha(x_0)$. In this case, we define μ as the suborder $[x_0, \alpha(x_0)[:= \{x \in \lambda \mid x_0 \leq x < \alpha(x_0)\}$. Because α preserves the order, we immediately get $\alpha^k(x_0) < \alpha^{k'}(x_0)$ for all integers $k < k'$, and even $\alpha^k(x) < \alpha^{k'}(x)$ for all $x \in \mu$ and all integers $k < k'$.

We consider now the subset X of all elements of λ of the form $\alpha^k(x)$, for some $x \in \mu$ and for some $k \in \mathbb{Z}$. By construction, it is isomorphic to $\mu \times \mathbb{Z}$, since for each $k \in \mathbb{Z}$, α^k is an isomorphism from $\mu = [x_0, \alpha(x_0)[$ to $[\alpha^k(x_0), \alpha^{k+1}(x_0)[$. Moreover, the convexity of X is deduced from the convexity of $[x_0, \alpha(x_0)[$.

Now, we prove the implication from $\neg$ii) to $\neg$i). We suppose that there exists some isomorphism ι from a convex subset X of λ to the linear order $\mu \times \mathbb{Z}$, with μ being a non-empty linear order. We define, for $x \in \lambda$, the element $\alpha(x)$ as being either x if

$x \notin X$, or $\iota^{-1}\big(\langle y, k{+}1\rangle\big)$ if $x \in X$ and $\iota(x) = \langle y, k\rangle$, with $y \in \mu$ and $k \in \mathbb{Z}$. The obtained function α is clearly an automorphism of λ, and it is non-trivial: $\alpha(x) \neq x$ for any $x \in X$. $\square$

In the next claim, we prove that, assuming the existence of a non-trivial automorphism for λ, it is impossible to define regularly a choice function or a well order over λ, and also to define all the elements of λ regularly.

The proofs of these implications are similar to each other, and are a direct corollary of Fact 1.5 on page 32, stating that if α is an automorphism of λ, then for any MSO[<] formula $\varphi(x_0, \ldots, x_{p-1}, X_0, \ldots, X_{q-1})$, we have $\lambda \vDash \varphi(x_0, \ldots, x_{p-1}, X_0, \ldots, X_{q-1})$ if and only if $\lambda \vDash \varphi\big(\alpha(x_0), \ldots, \alpha(x_{p-1}), \alpha(X_0), \ldots, \alpha(X_{q-1})\big)$.

Claim 4.8. *The condition:*

¬i) λ is non-rigid,

implies the following three:

¬iv) λ does not admit any regular choice function,

¬v) λ does not admit any regular well order,

¬vi) not every element of λ is regularly definable.

Proof. We suppose that λ admits a non-trivial automorphism α, and we can therefore consider an element $x_0 \in \lambda$ such that $\alpha(x_0) \neq x_0$.

Proof of ¬iv) We define X as the subset $\{\alpha^k(x_0) \mid k \in \mathbb{Z}\}$. If $\varphi(X, x)$ is an MSO[<] formula such that $\lambda \vDash \varphi\big(X, \alpha^k(x_0)\big)$ for some $k \in \mathbb{Z}$, then we also obtain $\lambda \vDash \varphi\big(\alpha(X), \alpha^{k+1}(x_0)\big)$, *i.e.* $\lambda \vDash \varphi\big(X, \alpha^{k+1}(x_0)\big)$: there cannot be a unique $x \in X$ such that $\lambda \vDash \varphi(X, x)$. This proves that λ cannot admit any regular choice function.

Proof of ¬v) Let $\varphi(x, y)$ be an MSO[<] formula. We suppose that it defines a linear order $\prec$ over λ, and we prove that said order cannot be a well order. Because $\prec$ is linear, we must have either $\lambda \vDash \varphi\big(x_0, \alpha(x_0)\big)$ or $\lambda \vDash \varphi\big(\alpha(x_0), x_0\big)$. In the former case, we have $\lambda \vDash \varphi\big(\alpha^k(x_0), \alpha^{k+1}(x_0)\big)$ for all $k \in \mathbb{Z}$, and in the latter case, we have $\lambda \vDash \varphi\big(\alpha^{k+1}(x_0), \alpha^k(x_0)\big)$ for all $k \in \mathbb{Z}$. In both cases, the subset $\{\alpha^k(x_0) \mid k \in \mathbb{Z}\}$ does not admit a least element by $\preceq$, which therefore cannot be a well order.

Proof of ¬vi) Let $\varphi(y)$ be an MSO[<] formula. If we suppose that $\lambda \vDash \varphi(x_0)$, then we have also $\lambda \vDash \varphi\big(\alpha(x_0)\big)$, and therefore φ cannot define the position x_0. This shows that not all elements of λ are regularly definable, and our proof is complete. $\square$

Finally, we prove that the existence of a regular choice function can be seen as a particular regular uniformisation of a regular relation:

Claim 4.9. *The condition:*

iii) every regular relation of λ admits a regular uniformisation

implies the following:

iv) λ admits a regular choice function.

Proof. Let $R \subseteq \binom{\{0,1\}}{\{0,1\}}^\lambda$ be the binary relation between λ-words over $\{0,1\}$ defined by: each pair $\binom{w}{\sigma}$ is in R if a unique position of σ is labelled by 1 and moreover this position is also labelled by 1 in w.

This way, the word w is identified with any non-empty subset X of λ, and σ selects a particular position of X.

A first thing to notice is that R is regular, since it is defined by the formula $\forall x.\ \neg\binom{0}{1}(x) \wedge \exists! x.\ \binom{1}{1}(x)$ (we recall that the definition of the existential quantifier with unicity can be found on page 32). Hence, if Item iii) is true, R admits a regular uniformisation F, which is *de facto* a choice function over λ: by selecting a particular $\sigma \in \{0,1\}^\lambda$ such that $\binom{w}{\sigma} \in R$, for each $w \in \{0,1\}^\lambda$ admitting at least one position labelled by 1, it selects a particular position among all the ones labelled by 1 in w.

Formally, if ψ defines the relation F, then the formula $\varphi(X, x)$ is defined as $x \in X \wedge \psi'$, where ψ' is obtained from ψ by substituting each of its atomic subformulae of the shape $\binom{1}{1}(y)$ by the formula $x = y$, each of its atomic subformulae of the shape $\binom{1}{0}(y)$ by the formula $y \in X \wedge x \neq y$, each of its atomic subformulae of the shape $\binom{0}{0}(y)$ by the formula $y \notin X$, and each of its atomic subformulae of the shape $\binom{0}{1}(y)$ by the formula $y \neq y$. $\square$

4.3 Tree decompositions and the possibility to define them regularly

After Section 4.2, we prove that the rigidity of a linear order λ implies the possibility to regularly uniformise the relations of domain λ, and also to define well orders and choice functions in MSO[<]. For this purpose, we develop a notion of *terms* and a notion of *condensation trees*. A *term* will be a finite representation of a finitary word w, while a *condensation tree* makes the link between the nodes of said term and w.

In Subsection 4.3.1, we define these two notions, while from Subsections 4.3.2 to 4.3.4, we introduce *tree decompositions*, a tool constructed from condensation trees, which we show to be related to the presence (or absence) of non-trivial automorphisms of a given word or linear order. Finally, in Subsection 4.3.5, we show how these tree decompositions can be defined in Monadic Second-Order Logic.

Before all of that, we notice that a finitary linear order constructed via the η-operation is necessarily non-rigid:

Fact 4.10. *If the construction of a finitary linear order λ involves the η-operation, then λ is non-rigid.*

Proof. Suppose that the construction of λ involves the η-operation: one of its convex subsets, name it X, is (isomorphic to) the linear order $\{\lambda_0, \dots, \lambda_{n-1}\}^\eta$, with each λ_i being a finitary linear order.

We recall the definition of $\{\lambda_0, \dots, \lambda_{n-1}\}^\eta$ (see page 27): we consider $u = \{0, \dots, n-1\}^\eta$, the unique (up to isomorphism) densely labelled word over the alphabet $\{0, \dots, n-1\}$ whose domain has no least nor greatest element (see Theorem 1.2 on page 24 and the definition of $\{0, \dots, n-1\}^\eta$ on page 26), and $\{\lambda_0, \dots, \lambda_{n-1}\}^\eta$ is obtained as $\Sigma_{q\in\mathrm{Dom}(u)}\lambda_{u(q)}$.

Now, we consider $v = u^{\mathbb{Z}}$, obtained as a bi-infinite number of copies of u. It is clear that $\mathrm{Dom}(v)$ is also without least nor greatest element, and that v is densely labelled. Therefore, there exists an isomorphism ι from v to u, and X is $\Sigma_{q\in\mathrm{Dom}(v)}\lambda_{u(\iota(q))}$. Since $\mathrm{Dom}(v)$ is $\mathrm{Dom}(u) \times \mathbb{Z}$, X is isomorphic to $\left(\Sigma_{q\in\mathrm{Dom}(u)}\lambda_{u(q)}\right) \times \mathbb{Z}$. Using Claim 4.7, we deduce that λ is non-rigid. $\square$

Thus, when we prove that if a linear order λ is rigid then it satisfies the regular-uniformisation property, we can assume without loss of generality that its construction does not involve the η-operation, meaning that it is scattered. We make this assumption when defining our notions of terms and condensation trees.

4.3.1 Terms and condensation trees

In this chapter, we define a *ranked symbol* as a symbol ℓ together with a (possibly empty) convex subset of $\mathbb{Z}$. Said convex subset is called the *arity* of ℓ, and is denoted by $\mathrm{ar}(\ell)$. We say that ℓ is *nullary* (resp. *unary*, *binary*) if $\mathrm{ar}(\ell) = \emptyset$ (resp. if $\mathrm{ar}(\ell) = \{0\}$, if $\mathrm{ar}(\ell) = \{1, 2\}$). A *ranked set* is a set of ranked symbols.

A *ranked tree* over a fixed ranked set S is defined inductively: if $\ell \in S$ is nullary, then there exists a ranked tree that is simply denoted by ℓ; if $\ell \in S$ is not nullary, and if for each $i \in \mathrm{ar}(\ell)$, t_i is a ranked tree, then there exists a ranked tree that is denoted $\ell[(t_i)_{i\in\mathrm{ar}(\ell)}]$. In a few cases, we use the following notations for said ranked tree: $\ell[t_0]$ when ℓ is unary, and $\ell[t_1, t_2]$ when ℓ is binary.

Each ranked tree $t = \ell[(t_i)_{i\in I}]$ can be seen as a structure consisting of a set of *nodes* (which are formally elements of $\mathbb{Z}^*$—finite words labelled by integers), defined inductively:

if t is the ranked tree ℓ, with ℓ being a nullary symbol, then $\mathrm{Nodes}(t) = \{\epsilon\}$, and if t is the ranked tree $\ell[(t_i)_{i\in \mathrm{ar}(\ell)}]$, with ℓ being a non-nullary ranked symbol, and each t_i being a ranked tree, then $\mathrm{Nodes}(t) = \{\epsilon\} \sqcup \bigsqcup_{i\in I}\{i\cdot u \mid u \in \mathrm{Nodes}(t_i)\}$. The node ϵ is called the *root* of t; if u is a node labelled by a symbol of arity I, then $u\cdot i$ is called a *child* of u, for each $i \in I$, and u is the *father* of each of its children $u\cdot i$. A *leaf* is a node that has no children—it must be labelled by a nullary symbol. By $\mathrm{Leaves}(t)$ we denote the set of all leaves of t. A node that is not a leaf is called *internal*.

If u and v are two nodes of t (internal nodes or leaves), we say that u is a *predecessor* of v, which we denote $u \preceq_{\mathrm{pred}} v$ if there exists a sequence $u_0, \dots, u_{k-1}$ of nodes of t such that $u_0 = u$, $u_{k-1} = v$, and for all $i < k-1$, u_{i+1} is a child of u_i. Notice that this notion of predecessors coincide exactly with the notion of prefixes when u and v are seen as words (see the definition on page 25). It is clear that the obtained relation, $\preceq_{\mathrm{pred}}$, is an order over $\mathrm{Nodes}(t)$, but it is *a priori* not linear. The inductive definition of ranked trees guarantees that for every node v, the set $\{u \in \mathrm{Nodes}(\tau) \mid u \preceq_{\mathrm{pred}} v\}$ of its predecessors is finite.

Since nodes of t are words over the alphabet $\mathbb{Z}$, which is linearly ordered, we can also provide $\mathrm{Nodes}(t)$ with the lexicographic order $\leq_{\mathrm{lex}}$, which is linear. Once again, we refer to page 25 for the definition. Notice that for all nodes u and v, $u \preceq_{\mathrm{pred}} v$ implies $u \leq_{\mathrm{lex}} v$.

Let $\mathbb{A}$ be an alphabet. We consider two families of ranked sets using the alphabet $\mathbb{A}$. First, a *term* over $\mathbb{A}$ is a ranked tree over the ranked set $\mathbb{A} \sqcup \{(\cdot), (\cdot)^{\omega}, (\cdot)^{\omega^\star}\}$ where each letter a of $\mathbb{A}$ is a nullary ranked symbol, $(\cdot)$ is a binary ranked symbol, and $(\cdot)^{\omega}$, $(\cdot)^{\omega^\star}$ are unary ranked symbols. Since the arities of all these ranked symbols are finite, a term is always a finite object.

We say that a word is *scattered* if its domain is, meaning that it does not contain any dense subset (see the definition on page 22). Each finitary scattered word over some alphabet $\mathbb{A}$ can be identified with a term τ over $\mathbb{A}$: the singleton word a can be identified with the term a; if w_1 and w_2 are two finitary words, identified with the terms τ_1 and τ_2 respectively, then the concatenation word $w_1\cdot w_2$ is identified with $(\cdot)[\tau_1, \tau_2]$; and if w is identified with the term τ, then the word w^{ω} (resp. $w^{\omega^\star}$) is identified with the term $(\cdot)^{\omega}[\tau]$ (resp. $(\cdot)^{\omega^\star}[\tau]$). When w is identified with τ, we also say that τ *represents*, or *is a representation of* w.

An example of a term can be seen on Figure 4.2: it represents the scattered finitary word $(a\cdot b)^{\omega}$. The labels are written under the nodes, and an arrow represents the father-child relation.

Second, a *condensation tree* over $\mathbb{A}$ is a ranked tree over the ranked set $\mathbb{A} \sqcup \{(\cdot), (\Sigma_{\omega}), (\Sigma_{\omega^\star})\}$, where, again, each letter a of $\mathbb{A}$ is a nullary ranked symbol and $(\cdot)$ is a binary ranked symbol, but (Σ_{ω}) and $(\Sigma_{\omega^\star})$ are ranked symbols of arity respectively $\omega = \{0, 1, 2, \dots\}$

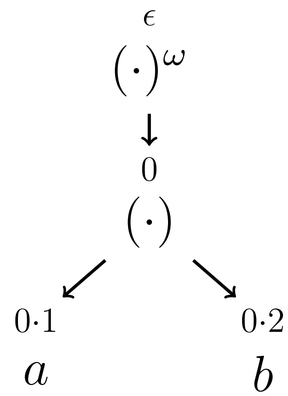


Figure 4.2: The term for the finitary word $(a \cdot b)^\omega$.

and $\omega^\star = \{\dots, -2, -1\}$. The leaves of a condensation tree, with the order $\leq_{\text{lex}}$, form a countable word over $\mathbb{A}$, which we denote by $\mathrm{Word}(t)$.

Each term τ over $\mathbb{A}$ corresponds to a particular condensation tree, which we denote by $\mathrm{Tree}(\tau)$: $\mathrm{Tree}(a)$ is the condensation tree a, for each $a \in \mathbb{A}$, $\mathrm{Tree}((\cdot)[\tau_1, \tau_2])$ is the condensation tree $(\cdot)[\mathrm{Tree}(\tau_1), \mathrm{Tree}(\tau_2)]$, $\mathrm{Tree}((\cdot)^\omega[\tau_0])$ is the condensation tree $(\Sigma_\omega)[(\mathrm{Tree}(\tau_0))_{i\in\omega}]$, and, symmetrically, $\mathrm{Tree}((\cdot)^{\omega^\star}[\tau_0])$ is the condensation tree $(\Sigma_{\omega^\star})[(\mathrm{Tree}(\tau_0))_{i\in\omega^\star}]$.

In this construction, each node v of $\mathrm{Tree}(\tau)$ is obtained from a particular node u of τ, and, very naturally, if τ is a representation of the scattered finitary word w, then the word $\mathrm{Word}(\mathrm{Tree}(\tau))$, which we simply write $\mathrm{Word}(\tau)$, is isomorphic to w. But on the other way, two distinct trees can represent isomorphic words: for instance, the words a^ω and $a \cdot (a^\omega)$ are isomorphic to each other, yet they are not represented by the same terms.

When we apply these rules to the term τ of Figure 4.3, we obtain the condensation tree on Figure 4.3, whose word $\mathrm{Word}(\tau)$ is (isomorphic to) $(a \cdot b)^\omega$.

As stated on page 24, words over a singleton alphabet are identified with their domains. Thus, we deduce a notion of terms and evaluation trees for linear orders when $\mathbb{A}$ is the singleton $\{\bullet\}$. Only, in this case, the ranked symbols $(\cdot)$, $(\cdot)^\omega$, and $(\cdot)^{\omega^\star}$ are respectively denoted by $(+)$, $(\times\omega)$, and $(\times\omega^\star)$, and we write $\mathrm{Order}(\tau)$ for $\mathrm{Word}(\tau)$. With this convention, Figure 4.4 depicts the condensation tree constructed from the term that represents the linear order $\omega + \omega^\star$.

4.3.2 Tree decompositions and how they relate to automorphisms

Here, we define a notion of *tree decomposition*, and explain how they relate to automorphisms of finitary words or linear orders. We make use of a notion of *condensations*: introduced

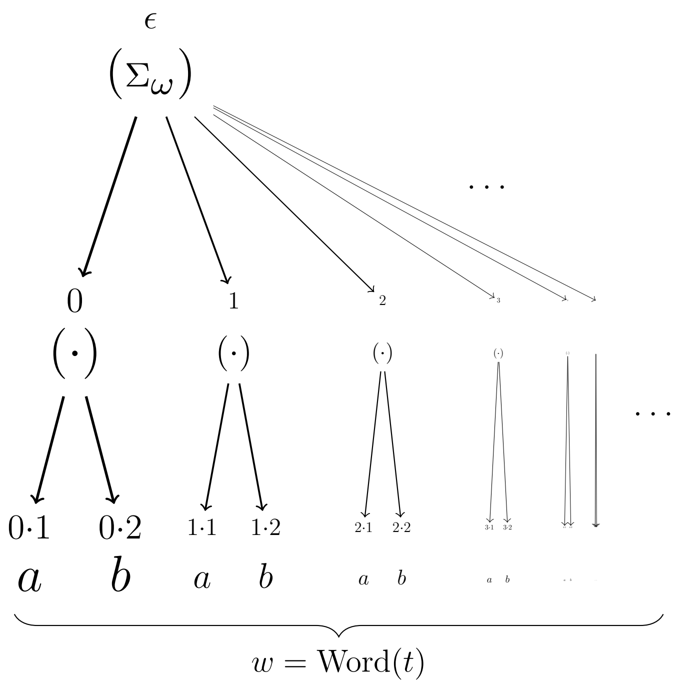


Figure 4.3: This condensation tree t is obtained from the term of Figure 4.2.

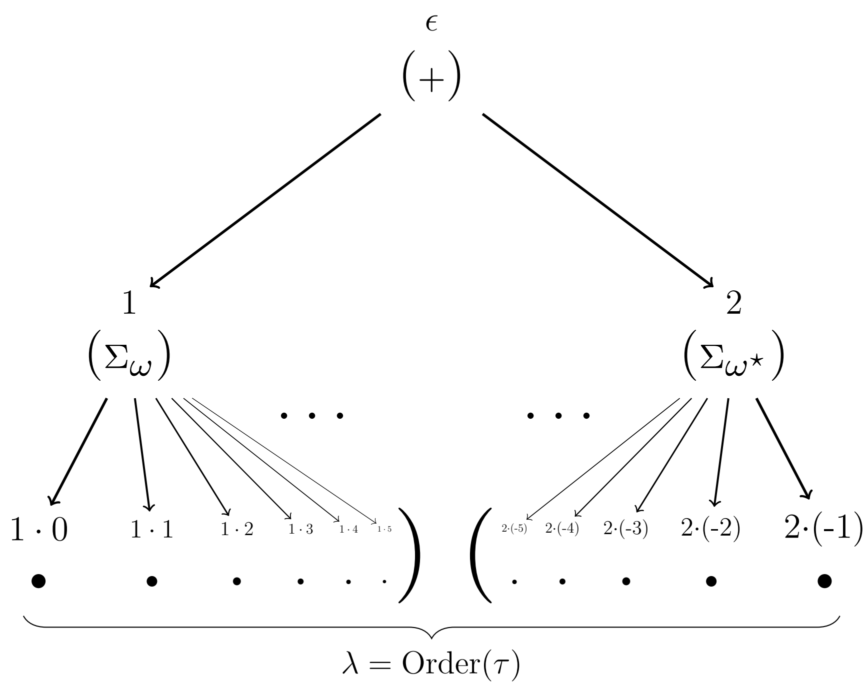


Figure 4.4: The condensation tree for the linear order $\omega + \omega^\star$.

in [CCP18], they are a very powerful tool to express properties over linear orders.

Let λ be a linear order and X be a subset of λ, not necessarily convex. We say that a subset U of λ is a *piece* of X if it is convex, included in X, and *maximal* in the following sense: if V is also a convex subset of λ included in X, then it is either disjoint from U or included in it. Every element of the subset X belongs to a particular piece of it, therefore, any subset X of λ is the disjoint union of all its pieces.

A *condensation* of a linear order λ is an equivalence relation C over a subset of λ, written $\mathrm{Dom}(C)$ and called the *domain* of C, such that every equivalence class of C is convex, meaning that for all $x \leq y \leq z$ in $\mathrm{Dom}(C)$ such that xCz, we also have xCy and yCz. Notice that, unlike the equivalence classes of C, the subset $\mathrm{Dom}(C)$ does not need to be convex. We call a *piece* of C any equivalence class of C.

Let τ be a term over an alphabet $\mathbb{A}$ and let w be a countable word over $\mathbb{A}$. A *tree decomposition with shape* τ of w is a family $\Xi = (C_u)_{u \in \mathrm{Nodes}(\tau)}$ of condensations of $\mathrm{Dom}(w)$, indexed by the nodes of τ, satisfying a certain number of conditions, for each $u \in \mathrm{Nodes}(\tau)$:

1) if u is ϵ, the root of the term, then C_u has $\mathrm{Dom}(w)$ for domain and is the full relation: xC_uy for all $x, y \in \mathrm{Dom}(w)$;
2) if u is a leaf, labelled by a letter $a \in \mathbb{A}$, then each piece of C_u must be a singleton $\{x\}$, with x being a position of w labelled by a;
3) if u is a binary internal node, labelled by $(\cdot)$, with children $u{\cdot}1 <_{\mathrm{lex}} u{\cdot}2$, then for every piece U of C_u, for each $i \in \{1, 2\}$, there exists a unique piece U_i of $C_{u \cdot i}$ that is included in U, and moreover we have $U_1 < U_2$, and $U_1 \sqcup U_2 = U$;
4) if u is a unary internal node, labelled by $(\cdot)^\omega$, with a unique child $u{\cdot}0$, then for every piece U of C_u, there exists an infinite number of pieces of $C_{u\cdot 0}$ contained in C_u, which we name $U_0, U_1, U_2, \ldots$ Moreover they satisfy $U_0 < U_1 < U_2 < \ldots$, $\bigsqcup_{i \in \omega} U_i = U$, and there is no other pieces of $C_{u \cdot 0}$ included in U;
5) in a symmetric way, if u is a unary internal node, labelled by $(\cdot)^{\omega^\star}$, with a unique child $u{\cdot}0$, then for every piece U of C_u, there exists an infinite number of pieces of $C_{u\cdot 0}$ that are contained in C_u, which we name $U_{-1}, U_{-2}, U_{-3}, \ldots$, moreover they satisfy $\ldots < U_{-3} < U_{-2} < U_{-1}$, $\bigsqcup_{i \in \omega^\star} U_i = U$, and there is no other pieces of $C_{u\cdot 0}$ included in U.

On Figure 4.5 is depicted a condensation tree of shape τ of the word $\mathrm{Word}(\tau)$, where τ is the term representing the word $(a \cdot b)^\omega$, depicted on Figure 4.2. It is actually the unique condensation tree of shape τ of this word. Each condensation is depicted with its pieces, bellow the leaves of $\mathrm{Tree}(\tau)$. Notice that, even though it is convenient to represent the condensations bellow the condensation tree $\mathrm{Tree}(\tau)$, they are indexed by the nodes of τ.

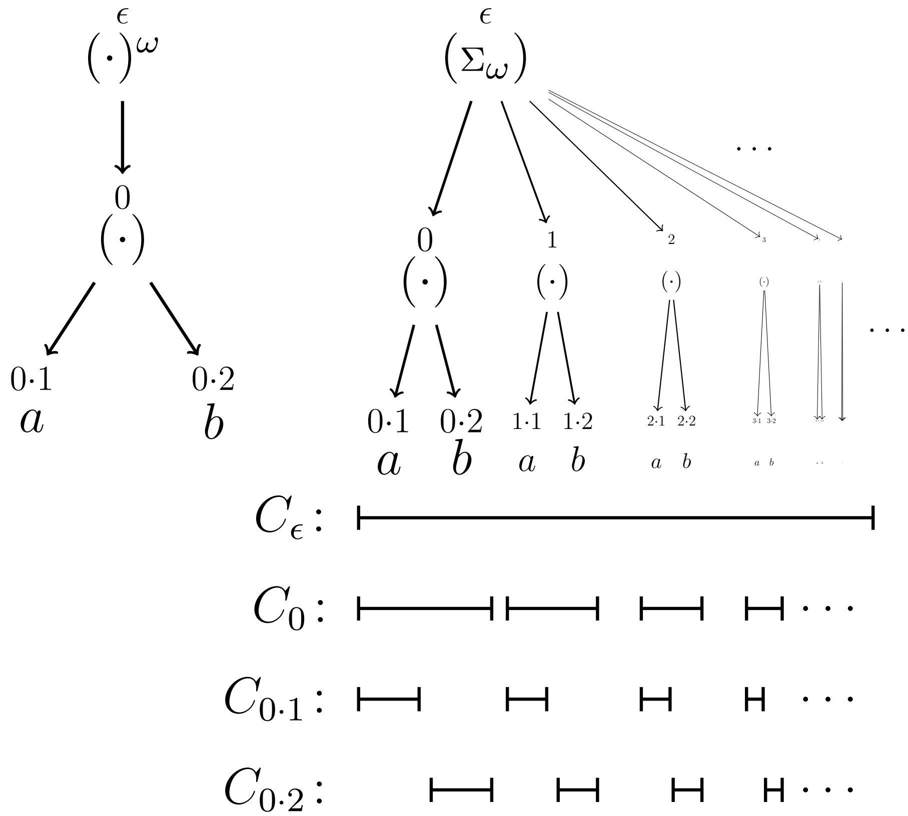


Figure 4.5: On the left-hand side of this picture is depicted the term τ representing the finitary word $(a \cdot b)^\omega$. On the right-hand side of this picture, we see the evaluation tree $\mathrm{Tree}(\tau)$ and a tree decomposition of $\mathrm{Word}(\tau)$ with shape τ.

A first thing to notice is that the propriety of being a tree decomposition is preserved by isomorphisms:

Lemma 4.11. *Let w and w' be two countable words over $\mathbb{A}$. If ι is an isomorphism from w to w' and if $\Xi = (C_u)_{u \in \mathrm{Nodes}(\tau)}$ is a tree decomposition with shape τ of w, then the family $\Xi' := \big(\iota(C_u)\big)_{u \in \mathrm{Nodes}(\tau)}$ is a tree decomposition with shape τ of w'.*

Proof. All the inductive conditions required for $\big(\iota(C_u)\big)_{u \in \mathrm{Nodes}(\tau)}$ to be a tree decomposition with shape τ, stated above, are preserved by isomorphism. $\square$

Proposition 4.12. *Let τ be a term over an alphabet $\mathbb{A}$ and $w \in \mathbb{A}^\circ$. There exists a bijection $\Xi \mapsto \iota_\Xi$ from the set of tree decompositions Ξ with shape τ of w, to the set of isomorphisms ι_Ξ from w to* $\mathrm{Word}(\tau)$.

In particular, words isomorphic to $\mathrm{Word}(\tau)$, and only them, admit a tree decomposition with shape τ.

The proof of Proposition 4.12 will be given in the next two subsections. Before that, let us observe the following corollary:

Corollary 4.13. *Let τ be a term over an alphabet $\mathbb{A}$ and w a countable word over $\mathbb{A}$ isomorphic to* $\mathrm{Word}(\tau)$*. Then w admits a unique tree decomposition with shape τ if and only if it is rigid.*

Proof. One implication is a consequence of the fact that if ι and ι' are two distinct isomorphisms from w to $\mathrm{Word}(\tau)$, then $\iota^{-1} \circ \iota'$ is a non-trivial automorphism of w.

The other implication is a consequence of the fact that if ι is an isomorphism from w to $\mathrm{Word}(\tau)$, and α a non-trivial automorphism of w, then $\iota \circ \alpha$ is a isomorphism from w to $\mathrm{Word}(\tau)$, distinct from ι. $\square$

4.3.3 From tree decompositions to isomorphisms

In this subsection, we explain how to construct the function $\Xi \mapsto \iota_\Xi$ mentioned in Proposition 4.12, and we prove that it is an injection.

If u is a node of a term τ, then it induces a *subterm* of τ: it is the set $\{v \in \mathrm{Nodes}(\tau) \mid u \preceq_{\mathrm{pred}} v\}$ of nodes of τ that admit u as a prefix. This term, which we write $\tau{\restriction}u$, naturally preserves the labels and the father-child relations, and we obtain from it a word $\mathrm{Word}(\tau{\restriction}u)$ that is isomorphic to a factor of $\mathrm{Word}(\tau)$.

With this notion, we can state and prove the following lemma:

Lemma 4.14. *Let $\Xi = (C_u)_{u \in \mathrm{Nodes}(\tau)}$ be a tree decomposition with shape τ of a word w, and let $u \in$* $\mathrm{Nodes}(\tau)$*. Then for every piece U of C_u, there exists an isomorphism $\iota_{\Xi,u,U}$ from $w{\restriction}_U$ to* $\mathrm{Word}(\tau{\restriction}u)$*.*

Proof. We construct the isomorphism $\iota_{\Xi,u,U}$ by induction on u.

- Let us suppose that u is a leaf of τ, labelled by a letter $a \in \mathbb{A}$. Then $\tau{\restriction}u$ is composed of this unique leaf, and $\mathrm{Word}(\tau{\restriction}u)$ is a word composed of a unique position labelled by a. By the definition of a tree decomposition, U, as a piece of C_u, is a singleton $\{x\}$, with $x \in \mathrm{Dom}(w)$ being labelled by a. Therefore, $w{\restriction}_U$ is clearly isomorphic to $\mathrm{Word}(\tau{\restriction}u)$.
- Let us suppose that u is an internal node of τ labelled by $(\cdot)$, having two children $u{\cdot}1 <_{\mathrm{lex}} u{\cdot}2$. By definition, U is $U_1 \sqcup U_2$, with $U_1 < U_2$, with each U_i being a piece of $C_{u \cdot i}$. Let us suppose that the lemma holds for $u{\cdot}1$ and $u{\cdot}2$, in order to show that it also holds for u: there exists an isomorphism $\iota_{\Xi,u\cdot 1,U_1}$ from $w{\restriction}_{U_1}$ to $\mathrm{Word}(\tau{\restriction}u{\cdot}1)$ and an isomorphism $\iota_{\Xi,u\cdot 2,U_2}$ from $w{\restriction}_{U_2}$ to $\mathrm{Word}(\tau{\restriction}u{\cdot}2)$.

By the definition of subterms, the word $\mathrm{Word}(\tau\restriction u{\cdot}1) \cdot \mathrm{Word}(\tau\restriction u{\cdot}2)$ is isomorphic to $\mathrm{Word}(\tau\restriction u)$. Let ι be an isomorphism from the former to the latter. Then, the function

$$\iota_{\Xi,u,U} = \left(\begin{array}{ccll} U & \to & \mathrm{Dom}(\mathrm{Word}(\tau\restriction u)) & \\ x & \mapsto & \begin{array}{ll} \iota\big(\langle \iota_{\Xi,u\cdot 1,U_1}(x), 1\rangle\big) & \text{if} \quad x \in U_1 \\ \iota\big(\langle \iota_{\Xi,u\cdot 2,U_2}(x), 2\rangle\big) & \text{if} \quad x \in U_2 \end{array} & \end{array} \right)$$

is clearly an isomorphism of words, and we have proven the lemma for the node u.

- Let us suppose now that u is an internal node of τ labelled by $(\cdot)^\omega$, having a unique child $u{\cdot}0$. By definition, U is $\bigsqcup_{i\in\omega} U_i$, with $U_0 < U_1 < U_2 < \dots$, with each U_i being a piece of $C_{u\cdot 0}$. Let us suppose that the lemma holds for $u{\cdot}0$: for each $i \in \omega$, there exists an isomorphism $\iota_{\Xi,u\cdot 0,U_i}$ from $w\restriction_{U_i}$ to $\mathrm{Word}(\tau\restriction u{\cdot}0)$.
  Again, it follows from the definition of the subterms that $\mathrm{Word}(\tau\restriction u)$ is isomorphic to $\mathrm{Word}(\tau\restriction u{\cdot}0)^\omega$. Let ι be the isomorphism from $\mathrm{Word}(\tau\restriction u{\cdot}0)^\omega$ to $\mathrm{Word}(\tau\restriction u)$.
  Now, for $x \in U$, let i be the unique natural number such that $x \in U_i$, and let y be $\iota_{\Xi,u\cdot 0,U_i}(x)$, in $\mathrm{Dom}(\mathrm{Word}(\tau\restriction u{\cdot}0))$, then we define $\iota_{\Xi,u,U}(x)$ as $\iota\big(\langle y, i\rangle\big)$, an element of $\mathrm{Dom}(\mathrm{Word}(\tau\restriction u))$. By construction, the obtained function $\iota_{\Xi,u,U}$ is an isomorphism from $w\restriction_U$ to $\mathrm{Word}(\tau\restriction u)$, which proves the lemma for the node u.
- We conclude by saying that the proof in the case when u is an internal node labelled by $(\cdot)^{\omega^\star}$ is completely symmetric to the previous case.

$\square$

Now that our isomorphisms $\iota_{\Xi,u,U}$ are constructed for all $u \in \mathrm{Nodes}(\tau)$ and all pieces U of C_u, we can define the wanted isomorphism ι_Ξ from w to $\mathrm{Word}(\tau)$ as being $\iota_{\Xi,\epsilon,\mathrm{Dom}(w)}$, since $\mathrm{Dom}(w)$ is the unique piece of the condensation C_ϵ.

A first thing to notice, which follows from our construction, is that distinct tree decompositions induce distinct isomorphisms:

Lemma 4.15. *If Ξ and Ξ' are two distinct tree decompositions with shape τ of a word w, then the isomorphisms ι_Ξ and $\iota_{\Xi'}$ are distinct.*

Proof. This lemma follows from the proof of Lemma 4.14: if u is a node of τ, and if the condensations C_u (from Ξ) and C'_u (from Ξ') differ, then it means that there exists a piece of C_u that is not a piece of C'_u (or the converse). Therefore, ι_Ξ, whose construction involves these pieces, differs from $\iota_{\Xi'}$. $\square$

Lemma 4.15 tells us that the function $\Xi \mapsto \iota_\Xi$ is an injection from the set of tree decompositions with shape τ of w to the set of isomorphisms from w to $\mathrm{Word}(\tau)$. In the next subsection, we show that it is also a surjection, and therefore a bijection.

4.3.4 From isomorphisms to tree decompositions

Let τ be a term over an alphabet $\mathbb{A}$. Recall that each node of $\mathrm{Tree}(\tau)$ is obtained from a unique node of τ. Let x, y be two positions of $\mathrm{Word}(\tau)$, meaning two leaves of $\mathrm{Tree}(\tau)$, and let $u \in \mathrm{Nodes}(\tau)$. We define that xC_uy if there exists a node v of $\mathrm{Tree}(\tau)$ obtained from u such that $v \preceq_{\mathrm{pred}} x$ and $v \preceq_{\mathrm{pred}} y$. The construction of $\mathrm{Tree}(\tau)$ ensures that two distinct nodes obtained from u are incomparable with respect to $\preceq_{\mathrm{pred}}$, hence, there can be at most one node v satisfying this property. This means that C_u is an equivalence relation, and moreover it is a condensation (because the set $\{x \in \mathrm{Leaves}(\mathrm{Nodes}(\tau)) \mid v \preceq_{\mathrm{pred}} x\}$ is convex for all $v \in \mathrm{Nodes}(\mathrm{Tree}(\tau))$). We can verify easily that the obtained tuple $(C_u)_{u\in\mathrm{Nodes}(\tau)}$ satisfies the inductive conditions described on page 117, therefore, it is a tree decomposition of $\mathrm{Word}(\tau)$ with shape τ. We call it the *canonical tree decomposition* of $\mathrm{Word}(\tau)$ with shape τ, and we denote it by Ξ_τ.

Claim 4.16. *The isomorphism ι_{Ξ_τ}, obtained from the canonical tree decomposition with shape τ, of* $\mathrm{Word}(\tau)$*, is* $\mathrm{id}_{\mathrm{Word}(\tau)}$*, the identity function of* $\mathrm{Word}(\tau)$*.*

Proof. We refer to the proof of Lemma 4.14: we can prove by induction that for every node $u \in \mathrm{Nodes}(\tau)$ and every piece U of C_u, the isomorphism $\iota_{\Xi_\tau,u,U}$ is the identity function of $\mathrm{Word}(\tau)\restriction_U$. Since ι_{Ξ_τ} is defined as $\iota_{\Xi_\tau,\epsilon,\mathrm{Dom}(\mathrm{Word}(\tau))}$, we conclude the claim. □

Now, we can prove that our function $\Xi \mapsto \iota_\Xi$ is a surjection:

Corollary 4.17. *Let ι be an isomorphism from a word $w \in \mathbb{A}^\circ$ to* $\mathrm{Word}(\tau)$*. Then there exists a tree decomposition Ξ with shape τ of w such that $\iota_\Xi = \iota$.*

Proof. Let $\Xi_\tau = (C_u)_{u\in\mathrm{Nodes}(\tau)}$ be the canonical tree decomposition of $\mathrm{Word}(\tau)$ with shape τ, and let Ξ be the family $\left(\iota^{-1}(C_u)\right)_{u\in\mathrm{Nodes}(\tau)}$. By Lemma 4.11, Ξ is a tree decomposition with shape τ of w. By the construction of Lemma 4.14, $\iota_\Xi = \iota \circ \iota_{\Xi_\tau}$. Since ι_{Ξ_τ} is the identity function of $\mathrm{Dom}(\mathrm{Word}(\tau))$, we have $\iota_\Xi = \iota$. □

With Lemma 4.15 and Corollary 4.17, we can conclude Proposition 4.12, telling that our construction $\Xi \mapsto \iota_\Xi$ is in fact a bijection from the set of tree decompositions with shape τ of w to the set of isomorphisms from w to $\mathrm{Word}(\tau)$.

4.3.5 Representing tree decompositions in Monadic Second-Order Logic

In this final subsection of Section 4.3, we prove a very important result: tree decompositions, our crucial tools, are definable in MSO[<].

First, we show how condensations can be represented in MSO[<]. Let X and D be two subsets of a linear order λ, with $X \subseteq D$. Let $x, y \in D$. For simplicity, by $[x, y]$ we denote the set of elements of λ between x and y, *i.e.* either the set $\{z \in \lambda \mid x \leq z \leq y\}$ (if $x \leq y$), or the set $\{z \in \lambda \mid y \leq z \leq x\}$ (if $y \leq x$). Now, we define that $x \sim_X^D y$ if $[x, y] \subseteq D$ and one of the following conditions is true: either $[x, y] \subseteq X$, or $[x, y] \cap X = \emptyset$.

Proposition 4.18. *The relation $\sim_X^D$ is a condensation of domain D.*

Proof. The fact that $\sim_X^D$ is reflexive and symmetric follows immediately from definition. We show that it is transitive: let $x \leq y \leq z$ in D, and suppose that $x \sim_X^D y \sim_X^D z$. A first thing to notice is that $[x, z] \subseteq D$, since $[x, y] \subseteq D$, $[y, z] \subseteq D$, and $[x, z] = [x, y] \cup [y, z]$. Then, we suppose that $[x, y] \subseteq X$ (the case $[x, y] \cap X = \emptyset$ being symmetric). Considering that $y \in X$, the condition $[y, z] \cap X = \emptyset$ does not hold, and therefore we necessarily have also $[y, z] \subseteq X$ (since $y \sim_X^D z$). Since $[x, z] = [x, y] \cup [y, z]$, we have $[x, z] \subseteq X$, and $x \sim_X^D z$. Therefore, $\sim_X^D$ is indeed an equivalence relation.

Now, it remains to show that the equivalence classes of $\sim_X^D$ are convex, which follows from definition. Indeed, if $x \leq y \leq z$ are elements of λ such that $x \sim_X^D z$, then, considering that $[x, y] \subseteq [x, z]$, we naturally have $[x, z] \subseteq D$, but also $[x, y] \subseteq X$ as long as $[x, z] \subseteq X$, and $[x, y] \cap X = \emptyset$ as long as $[x, z] \cap X = \emptyset$. Therefore, $x \sim_X^D y$, and the same for $y \sim_X^D z$. $\square$

Hence, any pair $\langle D, X\rangle$ of subsets of λ, with $X \subseteq D$, can be seen as a representation of a condensation over λ, by considering the relation $\sim_X^D$. Carton, Colcombet, and Puppis proved a reciprocal: all condensations of a linear order can be represented like this:

Lemma 4.19 ([CCP18, Lemma 35])**.** *For any condensation C of a linear order λ, there exist two subsets D and X of λ such that C is the relation $\sim_X^D$ over D.*

It shall be pointed as a remark that Lemma 35 of [CCP18] actually does not involve the subset D, since they assume condensations to be defined on the whole linear order λ. Yet, this result is easily generalisable to condensations defined on subsets of λ.

Also, notice that the representations are not unique, for if the pair $\langle D, X\rangle$ is a representation of the condensation C, then the pair $\langle D, D\backslash X\rangle$ is an example of a different representation for C.

For all subsets D, X, and for all positions x, y, the three conditions $[x, y] \subseteq D$, $[x, y] \subseteq X$, and $[x, y] \cap X = \emptyset$ are clearly definable by MSO[<] formulae, and therefore so is the condition $x \sim_X^D y$, definable by a formula $\varphi_{\text{cond}}(D, X, x, y)$. This tells that, if C is a condensation, we can express properties over C in MSO, by considering subsets D, X representing it. For convenience, we can even enrich our grammar by quantifiers $\exists^{\text{cond}}$ and $\forall^{\text{cond}}$, in order to quantify over condensations, and by an equality symbol $=^{\text{cond}}$, comparing condensations.

Having said that, we prove that also tree decompositions are expressible in MSO:

Proposition 4.20. *There exists an* $\mathrm{MSO}[\mathbb{A}, <]$ *formula* $\varphi^{\tau}_{\text{decomp}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}\big)$ *such that for any countable word* w *over* $\mathbb{A}$*, and any family* $(C_u)_{u\in\mathrm{Nodes}(\tau)}$ *of condensations of* $\mathrm{Dom}(w)$*, we have* $w \vDash \varphi^{\tau}_{\text{decomp}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}\big)$ *if and only if the family* $(C_v)_{v\in\mathrm{Nodes}(\tau)}$ *forms a tree decomposition with shape* τ *of* w*.*

To clarify, when we say that the formula $\varphi^{\tau}_{\text{decomp}}$ admits condensations as variables, then we formally implement them with second-order variables D_u and X_u. With that implementation, in $\varphi^{\tau}_{\text{decomp}}$, any atom of the shape xC_uy shall be understood as a shortened notation for the formula $\varphi_{\text{cond}}(D_u, X_u, x, y)$.

Proof. We make sure that all the items in the definition of a tree decomposition (see this definition on page 117) are definable in Monadic Second-Order Logic.

– The first item, for the node ϵ, is simple to express:

$$\varphi_{\text{root}}(C_\epsilon) := \forall x, y.\ xC_\epsilon y.$$

For the following items, we first define a formula $\varphi_{\text{piece}}(C, U)$ that expresses the fact that U, a subset of $\mathrm{Dom}(w)$, is a piece of the condensation C:

$$\begin{aligned}
\varphi_{\text{piece}}(C, U) :=\ & \\
& \varphi_{\text{convex}}(U) \wedge U{\neq}\emptyset \\
& \wedge \forall x{\in}U, y{\in}U.\ xCy \\
& \wedge \forall V.\ [\varphi_{\text{convex}}(V) \wedge \forall x{\in}V, y{\in}V.\ xCy] \\
& \qquad \implies U{\cap}V{=}\emptyset \vee V{\subseteq}U,
\end{aligned}$$

where $\varphi_{\text{convex}}(U)$ is naturally the formula $\forall x < y < z.\ (x{\in}U \wedge z{\in}U) \implies y{\in}U$, stating that U is a convex subset of $\mathrm{Dom}(w)$.

– Now, we can see that if u is a leaf of τ, labelled by some letter $a \in \mathbb{A}$, then the second item, applying to u, can be expressed by the formula:

$$\begin{aligned}&\varphi_a(C_u) := \\ &\qquad \forall U.\ \varphi_{\text{piece}}(C_u, U) \implies \exists x{\in}U.\ a(x) \wedge \forall y{\in}U.\ x = y.\end{aligned}$$

– If u is an internal node of τ labelled by $(\cdot)$ and having two children $u{\cdot}1 <_{\text{lex}} u{\cdot}2$, then the third item, applying to u, can be expressed by the formula:

$$\begin{aligned}&\varphi_{(\cdot)}(C_u, C_{u\cdot 1}, C_{u\cdot 2}) := \\ &\qquad \forall U.\ \varphi_{\text{piece}}(C_u,\, U) \implies \\ &\qquad\qquad \exists U_1,\, U_2.\ \ U_1{<}U_2 \\ &\qquad\qquad\qquad \wedge \quad U = U_1{\cup}U_2 \\ &\qquad\qquad\qquad \wedge \quad \forall V \subseteq U.\ \varphi_{\text{piece}}(C_{u\cdot 1}, V) \iff U_1 = V \\ &\qquad\qquad\qquad \wedge \quad \forall V \subseteq U.\ \varphi_{\text{piece}}(C_{u\cdot 2}, V) \iff U_2 = V.\end{aligned}$$

– Now, we express the condition for the case when u is an internal node of τ labelled by $(\cdot)^\omega$ and having one child $u{\cdot}0$. We recall that we must express in MSO$[\mathbb{A}, <]$ that, if U is a piece of C_u, then there exist infinitely many pieces $U_0 < U_1 < U_2 \ldots$ of $C_{u\cdot 0}$ such that $U = \bigsqcup_{n\in\omega} U_n$.
We consider two second-order variables U_{even} and U_{odd}, and our idea is to express that $U_{\text{even}} = \bigsqcup_{n\in\omega} U_{2n}$ and $U_{\text{odd}} = \bigsqcup_{n\in\omega} U_{2n+1}$, in a way that a piece of U_{even} or of U_{odd} is necessarily one of the U_n's.
To express this, it is sufficient to say that there exists a subset N of U that is isomorphic to ω and that intersects with all the pieces of U_{even} and of U_{odd} in only one element.
First, the fact that a subset N is isomorphic to ω can be defined by the formula:

$$\begin{aligned}&\varphi_{\text{iso},\omega}(N) := \\ &\qquad \forall x{\in}N.\ \exists y{\in}N.\ \varphi_{\text{succ}}(N, x, y) \\ &\qquad \wedge \exists x_0{\in}N.\ \ \forall x{\in}N.\ x_0 \ \ \leq x \\ &\qquad\qquad \wedge \forall P{\subseteq}N.\ \ \Big(x_0 \in N \\ &\qquad\qquad\qquad \wedge \big(\forall x, y{\in}N.\ (x{\in}P \wedge \varphi_{\text{succ}}(N, x, y)) \implies y{\in}P\big)\Big) \\ &\qquad\qquad\qquad \implies N \subseteq P,\end{aligned}$$

where $\varphi_{\text{succ}}(N,x,y)$ is the formula $x<y \wedge \forall z{\in}N.\ z{\leq}x \vee y{\leq}z$. The last three lines of the formula $\varphi_{\text{iso},\omega}(N)$ express the induction principle: N must be the smallest of its subsets that both contains x_0 and is stable by φ_{succ}.

Now, the fact that a subset V of U_{even} is a piece of it is defined by the formula:

$$\begin{aligned}
&\varphi_{\text{piece}}(U_{\text{even}},V) := \\
&\qquad\qquad V \subseteq U_{\text{even}} \\
&\qquad\qquad \wedge\, \varphi_{\text{convex}}(V) \\
&\qquad\qquad \wedge\, \forall W{\subseteq}U_{\text{even}}.\ \varphi_{\text{convex}}(W) \implies (V{\cap}W{=}\emptyset \vee W{\subseteq}V).
\end{aligned}$$

Notice that this is the second formula φ_{piece} which we define, but there is no confusion, since the first one, $\varphi_{\text{piece}}(C,U)$, had a condensation as one of its arguments, while this one has two subsets as argument.

Now, we have all we need to define in MSO[$\mathbb{A},<$] the full condition for the item:

$$\begin{aligned}
&\varphi_{(\cdot)^\omega}(C_u,C_{u\cdot 0}) := \\
&\qquad \forall U.\ \varphi_{\text{piece}}(C_u,U) \implies \\
&\qquad\quad \exists U_{\text{even}},U_{\text{odd}}.\ U_{\text{even}}{\cap}U_{\text{odd}}{=}\emptyset \wedge U_{\text{even}}{\cup}U_{\text{odd}}{=}U \\
&\qquad\quad \wedge\, \exists N{\subseteq}U.\ \ \varphi_{\text{iso},\omega}(N) \\
&\qquad\qquad\qquad \wedge\, \forall V.\quad \big(\varphi_{\text{piece}}(U_{\text{even}},V) \vee \varphi_{\text{piece}}(U_{\text{odd}},V)\big) \implies \\
&\qquad\qquad\qquad\qquad \big(\varphi_{\text{piece}}(C_{u\cdot 0},V) \\
&\qquad\qquad\qquad\qquad \wedge\, \exists x{\in}N{\cap}V.\ \forall y.\ y{\in}N{\cap}V \implies x{=}y\big).
\end{aligned}$$

– Finally, in the case when u is an internal node labelled by $(\cdot)^{\omega^\star}$, we obtain the wanted formula $\varphi_{(\cdot)^{\omega^\star}}(C_u,C_{u\cdot 0})$ with a symmetric construction.

Now that we defined our formulae corresponding to the different items, we can define our final formula:

$$\begin{aligned}\varphi^{\tau}_{\text{decomp}}\big((C_u)_{u\in\text{Nodes}(\tau)}\big) :=& \\ & \varphi_{\text{root}}(C_\epsilon) \\ & \wedge \bigwedge_{a\in\mathbb{A}} \bigwedge_{u\in\text{leaves}_a(\tau)} \varphi_a(C_u) \\ & \wedge \bigwedge_{u\in\text{Nodes}_{(\cdot)}(\tau)} \varphi_{(\cdot)}(C_u, C_{u\cdot 0}, C_{u\cdot 1}) \\ & \wedge \bigwedge_{u\in\text{Nodes}_{(\cdot)^\omega}(\tau)} \varphi_{(\cdot)^\omega}(C_u, C_{u\cdot 0}) \\ & \wedge \bigwedge_{u\in\text{Nodes}_{(\cdot)^{\omega^\star}}(\tau)} \varphi_{(\cdot)^{\omega^\star}}(C_u, C_{u\cdot 0}),\end{aligned}$$

where, $\text{leaves}_a(\tau)$, with $a \in \mathbb{A}$ is the set of leaves of τ labelled by a, and $\text{Nodes}_{(\cdot)}(\tau)$, $\text{Nodes}_{(\cdot)^\omega}(\tau)$, $\text{Nodes}_{(\cdot)^{\omega^\star}}(\tau)$ are the sets of nodes labelled by $(\cdot)$, $(\cdot)^\omega$, and $(\cdot)^{\omega^\star}$ respectively.

□

Now that we have defined in MSO[$\mathbb{A}$, <] this important property, we can use it to test if given finitary words are non-rigid:

Proposition 4.21. *Item i) of Theorem 4.1 is decidable: there exists an algorithm that inputs a term τ, representing a finitary word w over $\mathbb{A}$, and outputs* **YES** *if w is rigid, and* **NO** *if it is not.*

Proof. By Corollary 4.13, we know that w is rigid if and only if does not admit two distinct tree decompositions with shape τ. Using our formula $\varphi^{\tau}_{\text{decomp}}$ defined above, we can define this property in MSO[<]:

$$\begin{aligned}\neg\exists^{\text{cond}}(C_u)_{u\in\text{Nodes}(\tau)}.\ \exists^{\text{cond}}(C'_u)_{u\in\text{Nodes}(\tau)}.\ & \bigvee_{u\in\text{Nodes}(\tau)} \neg(C_u =^{\text{cond}} C'_u) \\ & \wedge\varphi^{\tau}_{\text{decomp}}\big((C_u)_{u\in\text{Nodes}(\tau)}\big) \\ & \wedge\varphi^{\tau}_{\text{decomp}}\big((C'_u)_{u\in\text{Nodes}(\tau)}\big).\end{aligned}$$

As we recalled in the introduction of this chapter, it is decidable whether w satisfies said formula, and we can conclude. □

Using this formula, we can also recognise words over a given finitary linear order:

Corollary 4.22. *If λ is a finitary scattered linear order, then the language of all countable words w over a given alphabet $\mathbb{A}$, such that* $\mathrm{Dom}(w)$ *is isomorphic to λ, is regular.*

Proof. The domain $\mathrm{Dom}(w)$ of a countable word w is isomorphic to λ if and only if it admits a tree decomposition with shape τ, with τ being a term over the singleton alphabet $\{\bullet\}$ which represents λ. In the construction of $\varphi^{\tau}_{\mathrm{decomp}}$, in the proof of Proposition 4.20, it suffices to replace the subformula $a(x)$, in the formula φ_a, by $\top$, to obtain an $\mathrm{MSO}[\mathbb{A},<]$ formula which expresses this property and which can be satisfied by w itself (instead of $\mathrm{Dom}(w)$). This suffices to prove the corollary. $\square$

4.4 Using evaluation trees to construct regular uniformisations

In this section, we finally prove that a rigid finitary linear order satisfies the regular-uniformisation property.

In Subsection 4.4.1, we explain that, in order to prove that such a linear order λ satisfies this property of regular uniformisations, it is sufficient to construct regular uniformisation to a restricted number of binary relations, over algebraic alphabets. Then, in Subsections 4.4.2 and 4.4.3, we conclude by proving that they indeed admit regular uniformisations. For this, we introduce a notion of τ-*evaluations*.

4.4.1 Algebraic formulation of the regular-uniformisation property

In this subsection, we prove that, in order to show that all regular relations over a given linear order admit regular uniformisations, it is enough to consider particular relations having $\circ$-semigroups as alphabets.

For this, we begin to explain how one can give a natural structure of $\circ$-semigroup to $\mathcal{P}(S)$, the set of subsets of a given $\circ$-semigroup S. Once again, we take this construction from [CCP18].

We denote by π_S its induced generalised product. We define, for each countable word Γ, over the alphabet $\mathcal{P}(S)$, $\pi_{\mathcal{P}(S)}(\Gamma)$ as the subset:

$$\left\{\pi_S(\gamma) \mid \gamma \in S^{\mathrm{Dom}(\Gamma)} \text{ and for all } x \in \mathrm{Dom}(\Gamma),\ \gamma(x) \in \Gamma(x)\right\}.$$

Carton, Colcombet, and Puppis showed that it is a generalised product over $\mathcal{P}(S)$. Hence, it induces a $\circ$-algebra $\langle \mathcal{P}(S), *, (\cdot)^\omega, (\cdot)^{\omega^\star}, (\cdot)^\kappa \rangle$. The first three functions are easily definable:

- for all $K_1, K_2 \subseteq S$, $K_1 * K_2$ is the set $\{s_1 * s_2 \mid s_1 \in K_1 \text{ and } s_2 \in K_2\}$;
- for each $K \subseteq S$, K^τ is the set $\{\pi_S(s_0 \cdot s_1 \cdot s_2 \cdots) \mid (s_i)_{i \in \omega} \in K^\omega\}$;
- symmetrically, $K^{\tau^\star}$ is the set $\{\pi_S(\cdots s_{-3} \cdot s_{-2} \cdot s_{-1}) \mid (s_i)_{i \in \omega^\star} \in K^{\omega^\star}\}$.

In the definition of K^τ, $s_0{\cdot}s_1{\cdot}s_2 \cdots$ stands for the ω-word over S where each position $i \in \omega$ is labelled by s_i. And symmetrically in the definition of $K^{\tau^\star}$.

A formal definition of $\{K_0, \ldots, K_{n-1}\}^\kappa$—with each K_i being a subset of S—can be found in [CCP18]. We do not give it here, since it is less intuitive, and we will not use it in our proofs.

Now, we prove the main lemma of this subsection. Let λ be a linear order, let $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^\lambda$ be a regular relation, which we will want to uniformise, and let S be a finite $\circ$-semigroup recognising R via some homomorphism h from $\binom{\mathbb{A}}{\mathbb{B}}^\circ$ to S. We set $H = h(R)$. Our lemma states that, in order to uniformise R, it is sufficient to uniformise a certain amount of relations over the algebraic product set $\binom{\mathcal{P}(S)}{S}$, treated as an alphabet:

Lemma 4.23. *Let us suppose that for every $s \in S$, there exists a regular uniformisation of the following relation:*

$$R_{\lambda,s} := \left\{ \binom{\Gamma}{\gamma} \in \binom{\mathcal{P}(S)}{S}^\lambda \mid \pi_S(\gamma) = s \text{ and for all } x \in \lambda,\ \gamma(x) \in \Gamma(x) \right\}.$$

Then the relation R admits a regular uniformisation. Moreover, such a regular uniformisation can be effectively constructed based on R and the uniformisations of the above relations.

Proof. Let us suppose that for each individual $s \in S$, the relation $R_{\lambda,s}$ admits a regular uniformisation $F_{\lambda,s}$, defined by an MSO$[\binom{\mathcal{P}(S)}{S}, <]$ formula $\psi^{\lambda,s}$. Before constructing an actual formula defining a uniformisation of R, we explain briefly how we proceed to assign, to each $w \in \Pi_{\mathbb{A}^\lambda}(R)$, a particular $\sigma \in \mathbb{B}^\lambda$ such that $\binom{w}{\sigma} \in R$.

First, we consider Γ the λ-word over the alphabet $\mathcal{P}(S)$ defined by $\Gamma(x) = \{h(\binom{w(x)}{b}) \mid b \in \mathbb{B}\}$ for each $x \in \lambda$. Then, we choose some $s \in H$ such that $\Gamma \in \Pi_{\mathcal{P}(S)^\lambda}(R_{\lambda,s})$ (because of the definition of Γ, of $\pi_{\mathcal{P}(S)}(\Gamma)$, and because $\Gamma \in \Pi_{\mathcal{P}(S)^\lambda}(R)$, we know that there necessarily exists such an element s). In the next step, $F_{\lambda,s}$, the uniformisation of $R_{\lambda,s}$, selects for us some $\gamma \in S^\lambda$: there exists a unique $\gamma \in S^\lambda$ such that $\binom{\Gamma}{\gamma} \in F_{\lambda,s}$. Finally, for each position $x \in \lambda$, we choose a particular letter $b \in \mathbb{B}$ such that $h(\binom{w(x)}{b}) = \gamma(x)$: once again, such a letter b necessarily exists since $\gamma(x) \in \Gamma(x)$. If we denote this letter by $\sigma(x)$, we obtain a word $\sigma \in \mathbb{B}^\lambda$ such that $h(\binom{w}{\sigma}) = s \in H$, and therefore such that $\binom{w}{\sigma} \in R$.

Now, we come back to these four steps and construct an MSO$[\binom{\mathbb{A}}{\mathbb{B}}, <]$ formula ψ that defines all these elements.

- First, we define an MSO$[\binom{\mathbb{A}}{\mathcal{P}(S)}, <]$ formula φ_1 such that for all $w \in \mathbb{A}^\lambda$, $\Gamma \in \mathcal{P}(S)^\lambda$, we have $\binom{w}{\Gamma} \vDash \varphi_1$ if and only if $\Gamma(x) = \{h(\binom{w(x)}{b}) \mid b \in \mathbb{B}\}$ for all $x \in \lambda$. Indeed, for all $a \in \mathbb{A}$, we can compute the set $K_a = \{h(\binom{a}{b}) \mid b \in \mathbb{B}\} \subseteq S$, and therefore we can define the formula:
$$\varphi_1 := \forall x. \bigvee_{a \in \mathbb{A}} \binom{a}{K_a}(x),$$
that fulfils the wanted condition.
- Now, we rename the letters of S: $S = \{s_0, \ldots, s_{|S|-1}\}$. This induces a natural linear order on S. Our point is to choose, for any word $\Gamma \in \mathcal{P}(S)^\lambda$, the least $s \in H$ such that $\binom{\Gamma}{\gamma} \in F_{\lambda,s}$ for some $\gamma \in S^\lambda$, if there exists one. First, remember that **MSO** is closed under projections, as stated in Proposition 1.47 on page 57. Hence, for each $s \in S$, we construct, from $\psi^{\lambda,s}$, a formula $\psi^{\lambda,s}_{\text{proj}} \in \text{MSO}[\mathcal{P}(S), <]$ such that a word $\Gamma \in \mathcal{P}(S)^\lambda$ satisfies it if and only if it belongs to $\Pi_{\mathcal{P}(S)^\lambda}(F_{\lambda,s})$ (which we recall is equal to $\Pi_{\mathcal{P}(S)^\lambda}(R_{\lambda,s})$).
Now, we want to define an MSO$[\binom{\mathcal{P}(S)}{S}, <]$ formula φ_2 such that for every product word $\binom{\Gamma}{\gamma}$, $\binom{\Gamma}{\gamma} \vDash \varphi_2$ if and only if there exists some $s_j \in H$ such that:
  - $\binom{\Gamma}{\gamma} \vDash \psi^{\lambda,s_j}$ (which in particular implies $\Gamma \vDash \psi^{\lambda,s_j}_{\text{proj}}$);
  - for every $i < j$ such that $s_i \in H$, $\Gamma \nvDash \psi^{\lambda,s_i}_{\text{proj}}$.

  The second condition can be expressed in MSO$[\binom{\mathcal{P}(S)}{S}, <]$, via the formulae $\psi^{\lambda,s}_{\text{proj}}$. Indeed, for every $s \in S$, if we substitute every atomic subformula of $\psi^{\lambda,s}_{\text{proj}}$ that is of the shape $K(x)$ by the formula $\bigvee_{t \in S} \binom{K}{t}(x)$, then we obtain an MSO$[\binom{\mathcal{P}(S)}{S}, <]$ formula ${\psi^{\lambda,s}_{\text{proj}}}'$ such that $\binom{\Gamma}{\gamma} \vDash {\psi^{\lambda,s}_{\text{proj}}}'$ if and only if $\Gamma \vDash \psi^{\lambda,s}_{\text{proj}}$.
Then, we can define the formula:
$$\varphi_2 := \bigvee_{j \in |S|,\ s_j \in H} \psi^{\lambda,s_j} \wedge \bigwedge_{i \in j,\ s_i \in H} \neg {\psi^{\lambda,s_j}_{\text{proj}}}'.$$
- In the next step, this is $\mathbb{B}$ we rename: $\mathbb{B} = \{b_0, \ldots, b_{|\mathbb{B}|-1}\}$, and we define E as the set $\{\langle a, s, b_j \rangle \in \mathbb{A} \times S \times \mathbb{B} \mid h(\binom{a}{b_j}) = s, \text{ and for all } i \in j,\ h(\binom{a}{b_i}) \neq s\}$. By construction, for all $x \in \lambda$, there exists a unique $b \in \mathbb{B}$ such that $\langle w(x), \gamma(x), b \rangle$ is in E, and, therefore, there exists a unique $\sigma \in \mathbb{B}^\lambda$ such that the triple satisfies the

following MSO$[\left(\begin{smallmatrix}\mathbb{A}\\S\\\mathbb{B}\end{smallmatrix}\right), <]$ formula:

$$\varphi_3 := \forall x. \bigvee_{<a,s,b>\in E} \left(\begin{smallmatrix}a\\s\\b\end{smallmatrix}\right)(x).$$

– Now that we have our three MSO formulae φ_1, φ_2, and φ_3, we construct our formula ψ, that defines a uniformisation of R. For this, we apply the same strategy as we did in Subsection 1.4.2: we use partitions. If $(X_K)_{K\subseteq S}$ and $(Y_s)_{s\in S}$ are respectively $2^{|S|}$- and $|S|$-partitions of λ, and if we define $\Gamma(X_\emptyset,\dots,X_S)$ and $\gamma(Y_{s_0},\dots,Y_{s_{|S|-1}})$ as the λ-words over respectively $\mathcal{P}(S)$ and S that are induced from these partitions, we can construct the following three MSO$[\left(\begin{smallmatrix}\mathbb{A}\\\mathbb{B}\end{smallmatrix}\right), <]$ formulae $\varphi'_1(X_\emptyset,\dots,X_S)$, $\varphi'_2(X_\emptyset,\dots,X_S,Y_{s_0},\dots,Y_{s_{|S|-1}})$, and $\varphi'_3(Y_0,\dots,Y_{|S|-1})$ such that for each pair $\left(\begin{smallmatrix}w\\\sigma\end{smallmatrix}\right) \in \left(\begin{smallmatrix}\mathbb{A}\\\mathbb{B}\end{smallmatrix}\right)$:
  i) $\left(\begin{smallmatrix}w\\\sigma\end{smallmatrix}\right) \vDash \varphi'_1(X_\emptyset,\dots,X_S)$ if and only if $\left(\begin{smallmatrix}w\\\Gamma(X_\emptyset,\dots,X_S)\end{smallmatrix}\right) \vDash \varphi_1$,
  ii) $\left(\begin{smallmatrix}w\\\sigma\end{smallmatrix}\right) \vDash \varphi'_2(X_\emptyset,\dots,X_S,Y_{s_0},\dots,Y_{s_{|S|-1}})$ if and only if $\left(\begin{smallmatrix}\Gamma(X_\emptyset,\dots,X_S)\\\gamma(Y_{s_0},\dots,Y_{s_{|S|-1}})\end{smallmatrix}\right) \vDash \varphi_2$,
  iii) $\left(\begin{smallmatrix}w\\\sigma\end{smallmatrix}\right) \vDash \varphi'_3(Y_{s_0},\dots,Y_{s_{|S|-1}})$ if and only if $\left(\begin{smallmatrix}w\\\gamma(Y_0,\dots,Y_{|S|-1})\\\sigma\end{smallmatrix}\right) \vDash \varphi_3$.

  Finally, we can give the final formula ψ:

$$\begin{aligned}\psi := \exists X_\emptyset,\dots,X_S,Y_0,\dots,Y_{|S|-1}.\ & \varphi^{2^{|S|}}_{\text{part}}(X_\emptyset,\dots,X_S) \wedge \varphi^{|S|}_{\text{part}}(Y_{s_0},\dots,Y_{s_{|S|-1}})\\ & \wedge \varphi'_1(X_\emptyset,\dots,X_S)\\ & \wedge \varphi'_2(X_\emptyset,\dots,X_S,Y_{s_0},\dots,Y_{s_{|S|-1}})\\ & \wedge \varphi'_3(Y_{s_0},\dots,Y_{s_{|S|-1}}).\end{aligned}$$

  By construction, ψ is such that for all $w \in \mathbb{A}^\lambda$, there exists a unique $\sigma \in \mathbb{B}^\lambda$ such that $\left(\begin{smallmatrix}w\\\sigma\end{smallmatrix}\right) \vDash \psi$, and in particular $\left(\begin{smallmatrix}w\\\sigma\end{smallmatrix}\right) \in R$: $\mathcal{L}^\lambda(\psi)$ is a regular uniformisation of the relation R.

□

Therefore, our aim now is to prove that, if λ is rigid, then each of these relations $R_{\lambda,s}$ admits a regular uniformisation, and our implication of Item iii) by Item i) of Theorem 4.1 will be proven. This is what we do in the next subsection.

4.4.2 Regularity of the relation R_s, in the particular cases $\{1,2\}$, ω, and $\omega^\star$

This section is devoted to the proof that the relation $R_{\lambda,s}$, defined in the previous subsection, is regular in the particular cases when λ being $\{1,2\}$, ω, and $\omega^\star$.

Proposition 4.24. *Let S be a finite $\circ$-semigroup. Then for any element $s \in S$, the three relations:*

$$R_{\{1,2\},s} := \left\{ \left(\begin{smallmatrix} K_1 \\ s_1 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} K_2 \\ s_2 \end{smallmatrix}\right) \mid s_1 \in K_1 \subseteq S,\ s_2 \in K_2 \subseteq S, \text{ and } s_1 * s_2 = s \right\},$$

$$R_{\omega,s} := \left\{ \left(\begin{smallmatrix} K_0 \\ s_0 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} K_1 \\ s_1 \end{smallmatrix}\right) \cdots \mid s_i \in K_i \subseteq S \text{ for all } i \in \omega, \text{ and } \pi(s_0 \cdot s_1 \cdots) = s \right\},$$

and

$$R_{\omega^\star,s} := \left\{ \cdots \left(\begin{smallmatrix} K_{-2} \\ s_{-2} \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} K_{-1} \\ s_{-1} \end{smallmatrix}\right) \mid s_i \in K_i \subseteq S \text{ for all } i \in \omega^\star, \text{ and } \pi(\cdots s_{-2} \cdot s_{-1}) = s \right\}$$

are all regular.

Proof.

- The relation $R_{\{1,2\},s}$ is finite, hence it is naturally regular. Here is a formula defining it:

$$\begin{aligned}
&\varphi^{\{1,2\},s} := \\
&\qquad \exists x_1, x_2.\ x_1 < x_2 \\
&\qquad\qquad \wedge \forall x.\ (x = x_1 \vee x = x_2) \\
&\qquad\qquad \wedge \bigvee_{\substack{s_1 \in K_1 \subseteq S \\ s_2 \in K_2 \subseteq S \\ s_1 * s_2 = s}} \left(\begin{smallmatrix} K_1 \\ s_1 \end{smallmatrix}\right)(x_1) \wedge \left(\begin{smallmatrix} K_2 \\ s_2 \end{smallmatrix}\right)(x_2).
\end{aligned}$$

  In this formula, x_1 is interpreted as the position 1 of $\{1,2\}$, and x_2 as the position 2.

- The relation $R_{\omega,s}$ is more interesting. Notice that it is enough for us to define a formula $\mu^{\omega,s} \in \mathrm{MSO}[S,<]$ such that for all ω-words γ over S, $\gamma \vDash \mu^{\omega,s}$ if and only if $\pi_S(\gamma) = s$. Indeed, if we do define such a formula, since we know how to add a coordinate in $\mathrm{MSO}[<]$, we can easily deduce a formula $\varphi^{\omega,s}$ defining $R_{\omega,s}$.

  Thus, let us define our formula $\mu^{\omega,s}$. For this, we make use of Ramsey's infinite theorem, which we stated on page 50. It tells us that $\gamma \in S^\omega$ is mapped to s by π_S if and only if there exists two elements $t, e \in S$ satisfying $t * e = t$, $e * e = e$, $t * e^\tau = s$, and γ admits a decomposition $\gamma = \gamma_0 \cdot \gamma_1 \cdot \gamma_2 \cdots$, where each γ_i is a finite non-empty word, such that $\pi_S(\gamma_i)$ is t if $i = 0$ and e if $i > 0$. We call such a decomposition a $\langle t, e\rangle$-*decomposition* of γ.

For each $t \in S$, we define a formula $\mu^t_{\text{finite}}(x_{\text{first}}, x_{\text{last}})$, x_{first} and x_{last} being two first-order variables, such that $\gamma \vDash \mu^t_{\text{finite}}(x_{\text{first}}, x_{\text{last}})$ if the subword $\gamma\restriction_{[x_{\text{first}}, x_{\text{last}}]}$, induced from the positions between x_{first} and x_{last}, is mapped to t by π_S. The idea is to introduce a second-order variable X_u for each $u \in S$: a position $x \in [x_{\text{first}}, x_{\text{last}}]$ will be in X_u if and only if $\pi_S\big(\gamma\restriction_{[x_{\text{first}}, x]}\big) = u$:

$$
\begin{aligned}
\mu^t_{\text{finite}}(x_{\text{first}}, x_{\text{last}}) :=&\\
&(\exists X_u)_{u\in S}.\\
&\quad \bigwedge_{u\in S} x_{\text{first}} \in X_u \iff u(x)\\
&\quad \wedge \forall x, y.\ \big(x_{\text{first}} \leq x < x_{\text{last}} \wedge \mathsf{s}(x,y)\big)\\
&\qquad\qquad \implies \bigwedge_{t_1,t_2\in S} x \in X_{t_1} \implies \big(t_2(y) \iff y \in X_{t_1 * t_2}\big)\\
&\quad \wedge x_{\text{last}} \in X_t,
\end{aligned}
$$

$\mathsf{s}(x, y)$ being the successor relation, which we met before, defined here with the order by the formula $x<y \wedge \forall z.\ z\leq x \vee y\leq z$.

Notice that this strategy is similar as the one we used in Subsection 1.3.1 when we showed that our variety $\mathbf{C}$ was able to express the recognisability by finite semigroups.

Now that we defined this formula μ^t_{finite}, we can simply express in MSO[<] the existence of a $\langle t, e\rangle$-partition: we express the existence of an infinite number of positions $0 = x_0 < x_1 < x_2 \ldots$ such that $\gamma \vDash \mu^t_{\text{finite}}(x_0, x_1{-}1)$, and $\gamma \vDash \mu^e_{\text{finite}}(x_n, x_{n+1}{-}1)$ for every natural number $n > 0$.

$$
\begin{aligned}
\mu^{t,e}_{\text{decomp}} :=&\\
&\exists x_0.\ \forall x.\ x_0 \leq x\\
&\quad \wedge \exists N.\ x_0 \in N \wedge \forall x{\in}N.\ \exists y{>}x.\ y \in N\\
&\quad \wedge \exists x_1{\in}N,\ y.\ \varphi_{\text{succ}}(N, x_0, x_1) \wedge \mathsf{s}(y, x_1) \wedge \mu^t_{\text{finite}}(x_0, y)\\
&\quad \wedge \forall x_n{\in}N.\ x_n \neq x_0 \implies\\
&\qquad\qquad \exists x_{n+1}{\in}N, y.\ \varphi_{\text{succ}}(N, x_n, x_{n+1}) \wedge \mathsf{s}(y, x_{n+1}) \wedge \mu^e_{\text{finite}}(x_n, y),
\end{aligned}
$$

where $\varphi_{\text{succ}}(N, x, y)$ is the formula $x{<}y \wedge \forall z{\in}N.\ z{\leq}x \vee y{\leq}z$, already defined in Subsection 4.3.5.

It shall be understood that, in this formula, x_n and x_{n+1} are two fixed variables: they are not indexed by some natural number n.

Finally, we define our formula $\mu^{\omega,s}$ as

$$\bigvee_{\langle t,e\rangle\in D_s} \mu^{t,e}_{\text{decomp}},$$

D_s being the set of pairs $\langle t, e\rangle \in S^2$ such that $e * e = e$, $t * e = t$, and $t * e^\tau = s$.

– By reversing the order in our formula $\varphi^{\omega,s}$, we obtain a formula $\varphi^{\omega^\star,s}$ defining the relation $R_{\omega^\star,s}$, and our proof is complete.

□

Since all these relations are regular, Theorems 1.48 and 1.49 tell us that they do admit regular uniformisations, which we denote by $F_{\{1,2\},s}$, $F_{\omega,s}$, and $F_{\omega^\star,s}$. We use them in the next section, to show that the relations $R_{\lambda,s}$ defined in Lemma 4.23 also admit regular uniformisations.

4.4.3 τ-evaluations and how to use them to uniformise

In this subsection, we finally prove that the relations $R_{\lambda,s}$, defined in Lemma 4.23, admit regular uniformisations, under the assumption that λ is rigid. For this, we define the important notion of τ-evaluations.

In this whole subsection, we consider a finitary linear order λ, represented by a term τ over $\{\bullet\}$. We suppose that λ is rigid. By Corollary 4.13, this means that there exists a unique tree decomposition $\Xi = (C_u)_{u\in\text{Nodes}(\tau)}$ for λ with shape τ, and that ι_Ξ is the unique isomorphism from λ to $\text{Order}(\tau)$. The unicity of Ξ is crucial, because our construction of a uniformisation of R works relatively to any tree decomposition of λ: it there were more than one, then our obtained relation would not be functional.

We consider also a finite $\circ$-algebra S, and denote by π_S the induced generalised product of it (see Theorem 1.27 on page 49).

Let f be a function from $\text{Nodes}(\text{Tree}(\tau))$ to S. We say that it is a τ-evaluation for S if it satisfies the following conditions:

– $f(u) = f(u{\cdot}1){*}f(u{\cdot}2)$ for every node u labelled by $(+)$, and having two children $u{\cdot}1 <_{\text{lex}} u{\cdot}2$;

– $f(u) = \pi_S\big(f(u{\cdot}0) \cdot f(u{\cdot}1) \cdot f(u{\cdot}2) \cdots\big)$ for each node u labelled by (Σ_ω) and having children $u{\cdot}0 <_{\text{lex}} u{\cdot}1 <_{\text{lex}} u{\cdot}2 <_{\text{lex}} \ldots$;
– $f(u) = \pi_S\big(\cdots f(u{\cdot}(\text{-}3)) \cdot f(u{\cdot}(\text{-}2)) \cdot f(u{\cdot}(\text{-}1))\big)$ for each node u labelled by $(\Sigma_{\omega^\star})$ and having children $\ldots <_{\text{lex}} u{\cdot}(\text{-}3) <_{\text{lex}} u{\cdot}(\text{-}2) <_{\text{lex}} u{\cdot}(\text{-}1)$.

In the second condition, $f(u{\cdot}0) \cdot f(u{\cdot}1) \cdot f(u{\cdot}2) \cdots$ denotes the ω-word over S where each position $i \in \omega$ is labelled by the element $f(u{\cdot}i) \in S$, and symmetrically in the third condition.

To each τ-evaluation f for S corresponds the word $\gamma \in S^\lambda$ defined by $\gamma(x) = f(\iota_\Xi(x))$ for each $x \in \lambda$ (remember that ι_Ξ is the unique isomorphism from λ to Order(τ) = Leaves(Tree(τ))). And reciprocally, to each word $\gamma \in S^\lambda$ corresponds the τ-evaluation f defined by $f(u) = \pi_S\big(\gamma{\restriction_U}\big)$, with U being the piece $\{v \in \text{Leaves}(\text{Tree}(\tau)) \mid u \preceq_{\text{pred}} v\}$ of C_u, for each $u \in$ Nodes(Tree(τ)). Hence, there is a natural one-to-one correspondence between λ-words over S and τ-evaluations.

In Subsection 4.4.1, we gave $\mathcal{P}(S)$, the set of subsets of S, a structure of $\circ$-semigroup. Thus, we also define a notion of τ-evaluations for $\mathcal{P}(S)$: those are the functions, from Nodes(Tree(τ)) to $\mathcal{P}(S)$, which satisfy the same conditions (with $\mathcal{P}(S)$ instead of S). Once again, there is a natural correspondence between λ-words over $\mathcal{P}(S)$ and τ-evaluations for $\mathcal{P}(S)$.

In particular, if f_S and $f_{\mathcal{P}(S)}$ are τ-evaluations, for $\mathcal{P}(S)$ and S respectively, and if, for every leaf u of Tree(τ), $f_S(u) \in f_{\mathcal{P}(S)}(u)$, then, by the definition of the generalised product $\pi_{\mathcal{P}(S)}$, we know that $f_S(u) \in f_{\mathcal{P}(S)}(u)$ for all nodes u, and in particular for the root, ϵ. For this reason, under this assumption, if Γ and γ are the λ-words corresponding to $f_{\mathcal{P}(S)}$ and f_S respectively, we must have $\binom{\Gamma}{\gamma} \in R_{\lambda, f_S(\epsilon)}$.

Let $s \in S$. Considering the previous paragraph, in order to define a uniformisation of $R_{\lambda,s}$, we have to choose, for each Γ such that $s \in \pi_{\mathcal{P}(S)}(\Gamma)$, a particular τ-evaluation f_S over S such that $f_S(\epsilon) = s$. This is what we do now: let Γ be a λ-word over $\mathcal{P}(S)$ such that $s \in \pi_{\mathcal{P}(S)}(\Gamma)$, and let $f_{\mathcal{P}(S)}$ be the τ-evaluation for $\mathcal{P}(S)$ corresponding to it. We say that f_S, a τ-evaluation for S, is *obtained* from $f_{\mathcal{P}(S)}$ and s if it satisfies the following conditions:

– $f_S(\epsilon)$ is equal to s.
– for every node u of Tree(τ) labelled by $(+)$ and having two children $u{\cdot}1 <_{\text{lex}} u{\cdot}2$, the uniformisation $F_{\{1,2\}, f_S(u)}$ maps the word $f_{\mathcal{P}(S)}(u{\cdot}1) \cdot f_{\mathcal{P}(S)}(u{\cdot}2) \in \mathcal{P}(S)^{\{1,2\}}$ to the word $f_S(u{\cdot}1) \cdot f_S(u{\cdot}2) \in S^{\{1,2\}}$;
– for every node u of Tree(τ) labelled by (Σ_ω), having children $u{\cdot}0 <_{\text{lex}} u{\cdot}1 <_{\text{lex}} u{\cdot}2 <_{\text{lex}} \ldots$, the uniformisation $F_{\omega, f_S(u)}$ maps the word $f_{\mathcal{P}(S)}(u{\cdot}0) \cdot f_{\mathcal{P}(S)}(u{\cdot}1) \cdot f_{\mathcal{P}(S)}(u{\cdot}2) \cdots \in$

$\mathcal{P}(S)^\omega$ to the word $f_S(u{\cdot}0) \cdot f_S(u{\cdot}1) \cdot f_S(u{\cdot}2) \cdots \in S^\omega$;

– symmetrically, for every node u of $\mathrm{Tree}(\tau)$ labelled by $(\Sigma_{\omega^\star})$ and having children $\ldots <_{\mathrm{lex}} u{\cdot}(\text{-}3) <_{\mathrm{lex}} u{\cdot}(\text{-}2) <_{\mathrm{lex}} u{\cdot}(\text{-}1)$, the uniformisation $F_{\omega^\star, f_S(u)}$ maps the word $\cdots f_{\mathcal{P}(S)}(u{\cdot}(\text{-}3)) \cdot f_{\mathcal{P}(S)}(u{\cdot}(\text{-}2)) \cdot f_{\mathcal{P}(S)}(u{\cdot}(\text{-}1)) \in \mathcal{P}(S)^{\omega^\star}$ to the word $\cdots f_S(u{\cdot}(\text{-}3)) \cdot f_S(u{\cdot}(\text{-}2)) \cdot f_S(u{\cdot}(\text{-}1)) \in S^{\omega^\star}$;

where $F_{\{1,2\},s}$, $F_{\omega,s}$, and $F_{\omega^\star}$ are the uniformisations defined for each $s \in S$ on page 133.

Since they are uniformisations, there cannot be two such f_S's. In that case, we also say, naturally, that the word $\gamma \in S^\lambda$ corresponding to f_S is *obtained* from Γ and s. If u is a node of τ, and U a piece of C_u, we also say that the subword $\gamma{\restriction_U}$ is obtained from $\Gamma{\restriction_U}$ and s, considering the subtree $\tau{\restriction_u}$.

Finally, we define $F_{\lambda,s}$ as the set of pairs $\binom{\Gamma}{\gamma} \in \binom{\mathcal{P}(S)}{S}$, with $s \in \pi_{\mathcal{P}(S)}(\Gamma)$, and γ being obtained from Γ and s.

For all the reasons stated above, $F_{\lambda,s}$ is a uniformisation of $R_{\lambda,s}$. The whole point of it is that, very conveniently, it is be definable in $\mathrm{MSO}[\binom{\mathcal{P}(S)}{S}, <]$:

Proposition 4.25. *The uniformisation $F_{\lambda,s}$ of $R_{\lambda,s}$ is definable in* $\mathrm{MSO}[\binom{\mathcal{P}(S)}{S}, <]$.

Proof. Our method is to define, for each node u of the term τ, for each $t \in S$, and for each $K \subseteq S$ such that $t \in K$, a formula $\varphi^{\tau,u,K,t}_{\text{obtained}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U\big)$ such that for every piece U of C_u, we have $\binom{\Gamma}{\gamma} \vDash \varphi^{\tau,u,K,t}_{\text{obtained}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U\big)$ if and only if $\pi_{\mathcal{P}(S)}(\Gamma{\restriction_U}) = K$ and $\gamma{\restriction_U}$ is obtained from $\Gamma{\restriction_U}$ and s.

We define this formula by induction on u.

– If u is a leaf, then any piece U of C_u must be a leaf labelled by $\binom{K}{t}$:

$$\varphi^{\tau,u,K,t}_{\text{obtained}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U\big) := \exists x.\ \tbinom{K}{t}(x) \wedge \forall y.\ \big(y \in U \iff x = y\big).$$

– Let us suppose that u is a $(+)$-node of τ, with two children $u{\cdot}1 <_{\mathrm{lex}} u{\cdot}2$, and let us suppose that we have already defined our formulae $\varphi^{\tau,u\cdot1,K_1,t_1}_{\text{obtained}}$ and $\varphi^{\tau,u\cdot2,K_2,t_2}_{\text{obtained}}$, for all $t_1 \in K_1 \subseteq S$, $t_2 \in K_2 \subseteq S$. Let U be a piece of C_u. By definition, U can be decomposed as the union of U_1 and U_2, with each U_i being a piece of $C_{u\cdot i}$.

Since $\binom{\mathcal{P}(S)}{S}$ is finite, then so is the language $F_{\{1,2\},t}$. It is recognised by the formula:

$$\begin{aligned}\psi^{\{1,2\},t} :=\ & \\ & \exists x_1, x_2.\ x_1 < x_2 \wedge \forall x.\ (x = x_1 \vee x = x_2) \\ & \qquad \wedge \bigvee_{\binom{K_1}{t_1}\cdot\binom{K_2}{t_2} \in F_{\{1,2\},t}} \tbinom{K_1}{t_1}(x_1) \wedge \tbinom{K_2}{t_2}(x_2).\end{aligned}$$

In this formula, which is similar to the formula defining the relation $R_{\{1,2\},t}$ of Proposition 4.24, x_1 is also interpreted as the position 1 of $\{1,2\}$, and x_2 as the position 2. From $\psi^{\{1,2\},t}$, we construct our formula $\varphi^{\tau,u,K,t}_{\text{obtained}}$, where x_1 is substituted to our piece U_1, and x_2 to U_2.

$$\begin{aligned}\varphi^{\tau,u,K,t}_{\text{obtained}}\big((C_u)_{u\in\text{Nodes}(\tau)},U\big) :=&\\ \exists U_1,\ U_2.\ U_1{<}U_2 \wedge U_1{\cup}U_2{=}\ &U\\ \wedge\, \varphi_{\text{piece}}(C_{u\cdot 1},U_1)\wedge\ &\varphi_{\text{piece}}(C_{u\cdot 2},U_2)\\ \wedge \bigvee_{\binom{K_1}{t_1}\cdot\binom{K_2}{t_2}\in F_{\{1,2\},t}} &\varphi^{\tau,u\cdot 1,K_1,t_1}_{\text{obtained}}\big((C_u)_{u\in\text{Nodes}(\tau)},U_1\big)\\ &\wedge\varphi^{\tau,u\cdot 2,K_2,t_2}_{\text{obtained}}\big((C_u)_{u\in\text{Nodes}(\tau)},U_2\big).\end{aligned}$$

- Now, let us suppose that u is a $(\times\omega)$-node, with one child $u\cdot 0$, and let us suppose that we have already defined our formula $\varphi^{\tau,u\cdot 0,K_0,t_0}_{\text{obtained}}$, for all $t_0\in K_0\subseteq S$. Let U be a piece of C_u: it is decomposed as $U=\bigsqcup_{i\in\omega}U_i$, where $U_0<U_1<U_2<\ldots$ are pieces of $C_{u\cdot 0}$ (and there is no other piece of $C_{u\cdot 0}$ included in U).

  In a similar way as in the previous item, we construct a new formula from $\psi^{\omega,s}$— a formula defining the regular uniformisation $F_{\omega,s}$ from page 133. We consider, for each first-order variable x, a second-order variable U_x. The idea is that if x is interpreted as $i\in\omega$, then U_x will be interpreted as the piece U_i. Similarly, for each second-order variable X, we consider another second-order variable V_X: if X is interpreted as a subset $I\subseteq\omega$, then V_X will be interpreted as $\bigsqcup_{i\in I}U_i$. We consider the following transformation $\varphi\mapsto\varphi'$ of formulae:
  - $x<y$ is transformed into $U_x<U_y$;
  - $x=y$ is transformed into $U_x=U_y$;
  - $x\in X$ is transformed into $U_x\subseteq V_X$;
  - $\binom{K_0}{t_0}(x)$, with $t_0\in K_0\subseteq S$, is transformed into $\varphi^{\tau,u\cdot 0,K_0,t_0}_{\text{obtained}}\big((C_u)_{u\in\text{Nodes}(\tau)},U_x\big)$,
  - $\exists x.\ \varphi$ is transformed into $\exists U_x{\subseteq}U.\ \varphi_{\text{piece}}(C_{u\cdot 0},U_x)\wedge\varphi'$, with φ' being the transformation of φ;
  - $\exists X.\ \varphi$ is transformed into

$$\begin{aligned}\exists V_X{\subseteq}U.\ \forall U_x{\subseteq}U.\ \varphi_{\text{piece}}(C_{u\cdot 0},U_x)\implies \big(U_x{\subseteq}V_X\vee U_x{\cap}V_X{=}\emptyset\big)\\ \wedge\,\varphi'\qquad\qquad\qquad\qquad\qquad\qquad\end{aligned}$$

– naturally, $\neg\varphi$ and $\varphi \vee \psi$ are respectively transformed into $\neg\varphi'$ and $\varphi' \vee \psi'$.

The obtained transformation $\psi^{\omega,t'}$ of $\psi^{\omega,t}$ is our formula $\varphi^{\tau,u,K,t}_{\text{obtained}}\big((C_u)_{u\in\text{Nodes}(\tau)}, U\big)$.

– In the case when u is a $(\times\omega^\star)$-node, our construction is symmetric.

Now that we have constructed our formulae, it suffices to define $\psi^{\lambda,s}$ as:

$$\begin{aligned}\exists^{\text{cond}}(C_u)_{u\in\text{Nodes}(\tau)}.\ & \varphi^{\tau}_{\text{decomp}}\big((C_u)_{u\in\text{Nodes}(\tau)}\big)\\ & \wedge \exists U.\ \varphi_{\text{piece}}(C_\epsilon, U)\\ & \wedge \bigvee_{\substack{K\subseteq S\\ s\in K}} \varphi^{\tau,\epsilon,K,s}_{\text{obtained}}\big((C_u)_{u\in\text{Nodes}(\tau)}, U\big),\end{aligned}$$

where $\varphi^{\tau}_{\text{decomp}}\big((C_u)_{u\in\text{Nodes}(\tau)}\big)$ is the formula defined in Subsection 4.3.5, expressing that the family $(C_u)_{u\in\text{Nodes}(\tau)}$ is a tree decomposition with shape τ. As stated above, λ admits a unique such tree decomposition and moreover λ is the unique piece of C_ϵ. Therefore, $\psi^{\lambda,s}$ defines $F_{\lambda,s}$. $\square$

Now that we have proved Proposition 4.25, we can use Lemma 4.23 of Subsection 4.4.1, and conclude this section with the most important proposition of this chapter:

Proposition 4.26. *A rigid finitary linear order satisfies the regular-uniformisation property.*

4.5 Constructing regular choice functions, regular well orders, and regularly defining positions using evaluation trees

In this section, we prove the last implications of Theorem 4.1: we prove that if a finitary linear order λ is rigid, then it admits a regular choice function, a regular well order, and that all its positions are definable in MSO[<].

4.5.1 Regular choice functions

This first subsection is devoted to the construction of a regular choice function for a finitary linear order without non-trivial automorphism. Notice that, in this case, a construction of such a regular choice function can be deduced from Claim 4.9 together with Proposition 4.26. Yet, the construction we provide here is more straight-forward.

Proposition 4.27. *Let λ be a finitary linear order. If λ is rigid, then we can construct an* MSO[<] *formula that defines a choice function over λ.*

Proof. Let τ be a term over the singleton alphabet $\{\bullet\}$ representing λ, meaning that $\mathrm{Order}(\tau)$ is isomorphic to λ, and let us suppose that the identity is the unique automorphism of λ. According to Corollary 4.13, there exists a unique tree decomposition $\Xi = (C_u)_{i\in\mathrm{Nodes}(\tau)}$ of λ with shape τ.

Our aim is to define, for each $u \in \mathrm{Nodes}(\tau)$, a formula $\varphi^{\tau,u}_{\text{choice}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U, X, x\big)$ such that, if U is a piece of C_u intersecting with X (meaning that $U \cap X \neq \emptyset$), then there exists a unique $x \in U \cap X$ such that $\lambda \vDash \varphi^{\tau,u}_{\text{choice}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U, X, x\big)$.

If we can do it, then it will be enough to define $\varphi^{\tau}_{\text{choice}}(X, x)$ as the formula:

$$\begin{aligned}\exists^{\text{cond}}(C_u)_{u\in\mathrm{Nodes}(\tau)}.\ \varphi^{\tau}_{\text{decomp}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}\big)&\\ \wedge \exists U.\ \varphi_{\text{piece}}(C_\epsilon, U)&\\ \wedge\, \varphi^{\tau,\epsilon}_{\text{choice}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U, X, x\big)&,\end{aligned}$$

to get our regular choice function.

Once again, we define these formulae inductively.

- If $u \in \mathrm{Leaves}(\tau)$, then a piece of C_u is a singleton. Therefore, it is sufficient to define the formula $\varphi^{\tau,u}_{\text{choice}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U, X, x\big)$ as $x{\in}U$.
- Let $u \in \mathrm{Nodes}(\tau)$, labelled by $(+)$, with two children $u{\cdot}1$ and $u{\cdot}2$, and let U be a piece of C_u that intersects with X. We suppose that we have defined our formulae $\varphi^{\tau,u\cdot 1}_{\text{choice}}$ and $\varphi^{\tau,u\cdot 2}_{\text{choice}}$. There exists a unique pair $\langle U_1, U_2\rangle$ of subsets of U such that each U_i is a piece of $C_{u\cdot i}$, $U_1 < U_2$, and $U_1 \sqcup U_2 = U$.

  Since U intersects with X, then so does U_1 or U_2. Knowing this, to select a particular element of $U \cap X$, we can select either a particular element of $U_1 \cap X$ when there exists one (using $\varphi^{\tau,u\cdot 1}_{\text{choice}}$), or a particular element of $U_2 \cap X$ in the opposite case.

  This procedure is naturally definable in MSO[<], and we obtain $\varphi^{\tau,u}_{\text{choice}}$ as the formula:

$$\begin{aligned}\exists U_1, U_2{\subseteq}U.\ \varphi_{\text{piece}}(C_{u\cdot 1}, U_1) \wedge \varphi_{\text{piece}}(C_{u\cdot 2}, U_2)&\\ \wedge \Big(U_1{\cap}X{\neq}\emptyset \implies \varphi^{\tau,u\cdot 1}_{\text{choice}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U_1, X, x\big)\Big)&\\ \wedge \Big(U_1{\cap}X{=}\emptyset \implies \varphi^{\tau,u\cdot 2}_{\text{choice}}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U_2, X, x\big)\Big)&.\end{aligned}$$

- Let $u \in \mathrm{Nodes}(\tau)$, labelled by $(\times\omega)$, with one child $u{\cdot}0$, and let U a piece of C_u intersecting with X. We suppose that we have defined our formula $\varphi^{\tau,u\cdot 0}_{\text{choice}}$. There exists a unique family $(U_i)_{i\in\omega}$ of pieces of $C_{u\cdot 0}$ included in U such that $U_0 < U_1 < U_2 < \dots$ and $\bigsqcup_{i\in\omega} U_i = U$.

Here, our strategy is to select the least i such that $U_i \cap X \neq \emptyset$, and to choose a particular element in this set $U_i \cap X$ via the formula $\varphi^{\tau,u\cdot 0}_{\text{choice}}$.

Here again, this procedure is definable in MSO, via the formula $\varphi^{\tau,u}_{\text{choice}}\big((C_u)_{u\in\text{Nodes}(\tau)}, U, X, x\big)$ defined as:

$$\begin{aligned}\exists V{\subseteq}U.\ & V{\cap}X{\neq}\emptyset \wedge \varphi_{\text{piece}}(C_{u\cdot 0}, V)\\ & \wedge \forall W.\ \big(W{\cap}X{\neq}\emptyset \wedge \varphi_{\text{piece}}(C_{u\cdot 0}, W)\big) \implies (V{=}W \vee V{<}W)\\ & \wedge \varphi^{\tau,u\cdot 0}_{\text{choice}}\big((C_u)_{u\in\text{Nodes}(\tau)}, V, X, x\big).\end{aligned}$$

– The construction of the formula when the node is labelled by $(\times\omega^\star)$ is symmetric, and we can conclude the proof of our proposition.

□

4.5.2 Regular well orders

In the second subsection, we prove that, if λ is a finitary linear order that satisfies the same crucial rigidity assumption, then it admits a well order that can be defined by an MSO[<] formula $\varphi^{\lambda}_{\text{wo}}(x, y)$.

For this, we define, for each finitary linear order, an *alternative lexicographic order*, which we will prove to be a well order.

First, we define $\sqsubseteq$, the *alternative order* of $\mathbb{Z}$, as follows: for $i, j \in \mathbb{Z}$, $i \sqsubseteq j$ if one of the three cases holds: $0 \le i \le j$, $j < 0 \le i$, or $j \le i < 0$. It is a well order, and it induces an alternative lexicographic order $\sqsubseteq_{\text{lex}}$ over $\mathbb{Z}^*$, the set of finite words over the infinite alphabet $\mathbb{Z}$. Unlike $\sqsubseteq$, $\sqsubseteq_{\text{lex}}$ is not a well order. Yet, it is when restricted to $\mathbb{Z}^{\le n}$, the set of finite words w over $\mathbb{Z}$ of length $|w| \le n$, as states the following proposition:

Proposition 4.28. *Let ϖ be a well order. Then, for all $n \in \omega$, $\varpi^{\le n}$, the set of finite words w over ϖ and of length $|w| \le n$, with its induced lexicographic order, is a well order.*

Proof. This is provable by induction on n: $\varpi^{\le 0}$ is $\{\epsilon\}$, which is obviously a well order.

Let us suppose now that $\varpi^{\le n}$ is a well order, and let us consider a non-empty subset X of $\varpi^{\le n+1}$, in order to show that it admits a least element.

If $\epsilon \in X$, then it is indeed the least element of X.

Let us assume the opposite case: $\epsilon \notin X$. Then we can write X as the disjoint union $\bigsqcup_{x\in\varpi} X_x$, where, for each element x in ϖ, X_x is the set $X \cap \{x \cdot u \mid u \in \varpi^{\le n}\}$.

Since $\epsilon \notin X$ and $X \neq \emptyset$, there must exist some x in ϖ such that $X_x \neq \emptyset$. Let x be the least of these elements (it necessarily exists, since ϖ is a well order). By the induction hypothesis, $\varpi^{\leq n}$ being a well order, $\{u \in \varpi^{\leq n} \mid x \cdot u \in X\}$ admits a least element u, and, by the definition of the lexicographic order, $x \cdot u$ is the least element of X_x. Also, if $y \cdot v$ is any word in $X \setminus X_x$, then we have $x < y$ by the very definition of x, and thus $x{\cdot}u <_{\text{lex}} y{\cdot}v$. Hence, $x \cdot u$ is the least element of X.

In both cases, we have shown that X has a least element, and we can conclude that $\varpi^{\leq n+1}$ is a well order. $\square$

Let τ be a term over the singleton alphabet $\{\bullet\}$. We recall that the leaves of $\mathrm{Tree}(\tau)$ are finite words over the alphabet $\mathbb{Z}$. Therefore, we can order them with $\sqsubseteq_{\text{lex}}^{\tau}$ defined as $\sqsubseteq_{\text{lex}} \restriction_{\mathrm{Order}(\tau)^2}$, where $\sqsubseteq_{\text{lex}}$ is the alternative lexicographic order of $\mathbb{Z}^*$, introduced above. Also, by construction, we can show that these words all have a length no greater than a certain *depth*[1]. Hence, $\sqsubseteq_{\text{lex}}^{\tau}$ is a well order over $\mathrm{Order}(\tau)$, considering Proposition 4.28.

If λ is a finitary scattered linear order, there exists some isomorphism ι from λ to $\mathrm{Order}(\tau)$, with τ being some term over $\{\bullet\}$, and we can naturally define $\sqsubseteq_{\text{lex}}^{\lambda,\iota}$, the *alternative lexicographic order* of λ with respect to ι, as $x \sqsubseteq_{\text{lex}}^{\lambda,\iota} y$ if $\iota(x) \sqsubseteq_{\text{lex}}^{\tau} \iota(y)$. Since $\sqsubseteq_{\text{lex}}^{\tau}$ is a well order, $\sqsubseteq_{\text{lex}}^{\lambda,\iota}$ also is one:

Proposition 4.29. *For every finitary linear order λ, represented by a term τ over $\{\bullet\}$, and every isomorphism from from λ to $\mathrm{Order}(\tau)$, the alternative lexicographic order $\sqsubseteq_{\text{lex}}^{\lambda,\iota}$ is a well order.*

Finally, now that we have defined a convenient well order for any finitary linear order, we show that it is definable in MSO[<] when λ is rigid:

Proposition 4.30. *Let λ be a finitary linear order. We suppose that λ is rigid. Then we can construct an* MSO[<] *formula defining a well order over λ.*

Proof. Let τ be a term representing λ. We know that there exists a unique tree decomposition $\Xi = (C_u)_{u \in \mathrm{Nodes}(\tau)}$ of λ with shape τ (once again, see Proposition 4.12).

Our aim is to construct, for each $u \in \mathrm{Nodes}(\tau)$, a formula $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U, x, y\big)$ such that for every piece U of C_u, and all $x, y \in U$, $\lambda \vDash \varphi_{\text{lex}}^{\lambda,\iota_\Xi,u}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U, x, y\big)$ if and only if $x \sqsubseteq_{\text{lex}}^{\lambda,\iota_\Xi} y$.

– If u is a leaf, then U is a singleton, and we define $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, U, x, y\big)$ as $\top$.

[1]The depth of a ranked tree being naturally defined by $\mathrm{depth}(\ell) = 0$ for nullary symbols ℓ, and $\mathrm{depth}(\ell[(t_i)_{i\in\mathrm{ar}(\ell)}]) = \max\big(\mathrm{depth}(t_i)\big)_{i\in\mathrm{ar}(\ell)} + 1$ for non-nullary symbols ℓ.

– Let u be a node of τ labelled by $(+)$, with two children $u{\cdot}1$ and $u{\cdot}2$. We suppose that we have defined the two relative formulae $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u\cdot 1}$ and $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u\cdot 2}$.

  Let U be a piece of C_u, we know that there exists a unique pair $\langle U_1, U_2\rangle$ of subsets of U such that each U_i is a piece of $C_{u\cdot i}$, $U_1 < U_2$, and $U_1 \sqcup U_2 = U$. Let $x, y \in U$. By the definition of $\sqsubseteq_{\text{lex}}^{\lambda,\iota_\Xi}$, if x and y are not in the same U_i, then $x \sqsubseteq_{\text{lex}}^{\lambda,\Xi} y$ if and only if $x \in U_1$ and $y \in U_2$. Hence, we can define $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u}\big((C_u)_{u\in\text{Nodes}(\tau)}, U, x, y\big)$ as the formula:

$$\begin{aligned}\exists U_1, U_2 \subseteq U.\ & \varphi_{\text{piece}}(C_{u\cdot 1}, U_1) \wedge \varphi_{\text{piece}}(C_{u\cdot 2}, U_2) \wedge \\ & \Big[\big(x \in U_1 \wedge y \in U_2\big) \\ & \vee \Big(x \in U_1 \wedge y \in U_1 \wedge \varphi_{\text{lex}}^{\lambda,\iota_\Xi,u\cdot 1}\big((C_u)_{u\in\text{Nodes}(\tau)}, U_1, x, y\big)\Big) \\ & \vee \Big(x \in U_2 \wedge y \in U_2 \wedge \varphi_{\text{lex}}^{\lambda,\iota_\Xi,u\cdot 2}\big((C_u)_{u\in\text{Nodes}(\tau)}, U_2, x, y\big)\Big)\Big]\end{aligned}$$

– Now, let u be a node of τ labelled by $(\times\omega)$, and having one child $u{\cdot}0$. We suppose to have defined our formula $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u\cdot 0}$. Let U be some piece of C_u. We know that there exists a infinite number of pieces of $C_{u\cdot 0}$, $U_0 < U_1 < U_2 \ldots$ such that $\sqcup_{i\in\omega} U_i$, and there exists no other piece of $C_{u\cdot 0}$ in U. Let $x, y \in U$: x is in some U_i, and y in some U_j. Here again, by the definition of $\sqsubseteq_{\text{lex}}^{\lambda,\iota_\Xi}$, if $i \neq j$, then $x \sqsubseteq_{\text{lex}}^{\lambda,\iota_\Xi} y$ if and only if $i < j$. Hence, we can define $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u}$ as follows:

$$\begin{aligned}\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u}\big((C_u)_{u\in\text{Nodes}(\tau)}, U, x, y\big) :=\ & \\ & \exists U_i \subseteq U, U_j \subseteq U.\ \varphi_{\text{piece}}(C_{u\cdot 0}, U_i) \wedge \varphi_{\text{piece}}(C_{u\cdot 0}, U_j) \\ & \wedge x \in U_i \wedge y \in U_j \\ & \wedge \quad \Big[U_i < U_j \\ & \qquad \vee \Big(U_i = U_j \wedge \varphi_{\text{lex}}^{\lambda,\iota_\Xi,u\cdot 0}\big((C_u)_{u\in\text{Nodes}(\tau)}, U_i, x, y\big)\Big)\Big].\end{aligned}$$

– The case when u is a node of τ labelled by $(\times\omega^\star)$ is completely symmetric to the previous case, only the subformula $U_i < U_j$ is replaced by $U_j < U_i$.

Now that our formulae $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,u}$ are defined, it suffices to define $\varphi_{\text{lex}}^{\lambda,\iota_\Xi,\iota_\Xi}(x, y)$ as the formula:

$$\begin{aligned}\exists (C_u)_{u\in\text{Nodes}(\tau)}, U.\ & \varphi_{\text{decomp}}^{\tau}\big((C_u)_{u\in\text{Nodes}(\tau)}\big) \\ & \wedge \varphi_{\text{piece}}(C_\epsilon, U) \\ & \wedge \varphi_{\text{lex}}^{\lambda,\iota_\Xi,\epsilon}\big((C_u)_{u\in\text{Nodes}(\tau)}, U, x, y\big).\end{aligned}$$

Since, as said previously, $\Xi = (C_u)_{u \in \mathrm{Nodes}(\tau)}$ is the *unique* tree decomposition of λ with shape τ, this formula defines $\sqsubseteq_{\mathrm{lex}}^{\lambda,\iota_\Xi}$, which we have proved to be a well order over λ. $\square$

4.5.3 Regularly defining the positions

In this third and final subsection, we prove that, under the same hypothesis about a linear order λ, we can construct, for each position x of λ, an MSO[<] formula $\varphi_{\mathrm{def}}^{\lambda,x}(y)$ that is satisfied for $y = x$ and only for $y = x$.

Proposition 4.31. *Let λ be a finitary linear order. We suppose that λ is rigid. Then, for every position x of it[2], we can construct a formula that defines it.*

Proof. Here again, let τ be a term over the singleton alphabet $\{\bullet\}$, representing λ, and let us suppose that λ is rigid. Let $\Xi = (C_u)_{u \in \mathrm{Nodes}(\tau)}$ be the unique tree decomposition of λ with shape τ, and ι_Ξ the unique isomorphism from λ to $\mathrm{Order}(\tau)$. We refer to the proof of Proposition 4.12 for the relation between the two of them.

We want to construct, for any leaf v of $\mathrm{Tree}(\tau)$, a formula $\varphi_{\mathrm{def}}^{\tau,v}(x)$ such that for every position x of λ, we have $\lambda \vDash \varphi_{\mathrm{def}}^{\tau,v}\big((C_u)_{u \in \mathrm{Nodes}(\tau)}, x\big)$ if and only $x = \iota_\Xi^{-1}(v)$.

To do this, we define, for each node u of $\mathrm{Tree}(\tau)$, a formula $\varphi_{\mathrm{def}}^{\tau,v}\big((C_u)_{u \in \mathrm{Nodes}(\tau)}, x\big)$ such that for all $x \in \lambda$, $\lambda \vDash \varphi_{\mathrm{def}}^{\tau,v}\big((C_u)_{u \in \mathrm{Nodes}(\tau)}, x\big)$ if and only if x is in the set $\{\iota_\Xi^{-1}(v') \mid v' \in \mathrm{Nodes}(\mathrm{Tree}(\tau))$ and $v \preceq_{\mathrm{pred}} v'\}$. Notice that when v is a leaf, the condition is exactly the same as the one we are aiming to prove (since v is the only leaf v' such that $v \preceq_{\mathrm{pred}} v'$).

As usually in this chapter, we construct our formula inductively. This time from top (the root) to bottom (the leaves).

- If v is the root ϵ, then all positions of λ are in the set $\{\iota_\Xi^{-1}(v') \mid v' \in \mathrm{Nodes}(\mathrm{Tree}(\tau))$ and $v \preceq_{\mathrm{pred}} v'\}$. Therefore, $\varphi_{\mathrm{def}}^{\tau,\epsilon}\big((C_u)_{u \in \mathrm{Nodes}(\tau)}, x\big) := \top$ satisfies our need.
- Let us suppose that we have constructed the formula $\varphi_{\mathrm{def}}^{\tau,v}$ for a node v of $\mathrm{Tree}(\tau)$ labelled by $(+)$, and let us construct the two formulae $\varphi_{\mathrm{def}}^{\tau,v\cdot 1}$ and $\varphi_{\mathrm{def}}^{\tau,v\cdot 2}$. Via the construction of $\mathrm{Tree}(\tau)$ from τ (which can be found on page 115), v is obtained from a node u of τ that is labelled by $(+)$.

  The set $\big\{\iota_\Xi^{-1}(v') \mid v' \in \mathrm{Nodes}(\mathrm{Tree}(\tau))$ and $v \preceq_{\mathrm{pred}} v'\big\}$, is some piece U of the condensation C_u. By definition, U can be divided into two disjoint subsets U_1 and U_2, with $U_1 < U_2$ and each U_i being a piece of the condensation $C_{u\cdot i}$, and there is no other piece of $C_{u\cdot i}$ included in U.

[2] In a procedure, while the linear order λ can naturally be given via a term, the position x can be given as a sequence of integers that position it in the condensation tree. More details about it can be found in Appendix C.

Therefore, the set $\{\iota_\Xi^{-1}(v') \mid v' \in \mathrm{Nodes}(\mathrm{Tree}(\tau)) \text{ and } v{\cdot}1 \preceq_{\mathrm{pred}} v'\}$ (resp. the set $\{\iota_\Xi^{-1}(v') \mid v' \in \mathrm{Nodes}(\mathrm{Tree}(\tau)) \text{ and } v{\cdot}2 \preceq_{\mathrm{pred}} v'\}$) is exactly U_1 (resp. U_2), and we can define $\varphi_{\mathrm{def}}^{\tau,v\cdot 1}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, x\big)$ as:

$$\begin{aligned}
&\varphi_{\mathrm{def}}^{\tau,v}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, x\big)\\
\wedge &\exists U_1.\ \varphi_{\mathrm{piece}}(C_{v\cdot 1}, U_1)\\
&\wedge x \in U_1,
\end{aligned}$$

and $\varphi_{\mathrm{def}}^{\tau,v\cdot 2}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, x\big)$ as the formula:

$$\begin{aligned}
&\varphi_{\mathrm{def}}^{\tau,v}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, x\big)\\
\wedge &\exists U_2.\ \varphi_{\mathrm{piece}}(C_{v\cdot 2}, U_2)\\
&\wedge x \in U_2,
\end{aligned}$$

– Finally, we treat the case when u is a node labelled by (Σ_ω) and has an infinite number of children $u\cdot 0 <_{\mathrm{lex}} u\cdot 1 <_{\mathrm{lex}} u\cdot 2 <_{\mathrm{lex}} \dots$ (here again, the case when it is labelled by $(\Sigma_{\omega^\star})$ is symmetric). The node v is obtained from a node u of τ, labelled by $(\times\omega)$.
Here, the set $\{\iota_\Xi^{-1}(v) \mid v \in \mathrm{Nodes}(\mathrm{Tree}(\tau)) \text{ and } u \preceq_{\mathrm{pred}} v\}$ is a piece U of the condensation C_u that is divided into an infinite number of pieces of $C_{u\cdot 0}$: $U = \bigsqcup_{i\in\omega} U_i$, with $U_0 < U_1 < U_2 < \dots$ There is no other piece of $C_{u\cdot 0}$ included in U.
This is why, for each $i \in \omega$, the set $\{\iota_\Xi^{-1}(v') \mid v' \in \mathrm{Nodes}(\mathrm{Tree}(\tau)) \text{ and } v{\cdot}i \preceq_{\mathrm{pred}} v'\}$ is exactly U_i, and it can be defined in MSO as the i-th least piece of $C_{u\cdot 0}$ included in U:

$$\begin{aligned}
\varphi_{\mathrm{def}}^{\tau,v\cdot i}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, x\big) := {}& \varphi_{\mathrm{def}}^{\tau,v\cdot i}\big((C_u)_{u\in\mathrm{Nodes}(\tau)}, x\big)\\
&\wedge \exists U_0 \subseteq U, \dots, U_i \subseteq U.\ \bigwedge_{0\le j\le i} \varphi_{\mathrm{piece}}(C_{u\cdot 0}, U_i)\\
&\wedge \bigwedge_{0\le j<i} U_j < U_{j+1}\\
&\wedge \forall V \subseteq U.\ \varphi_{\mathrm{piece}}(C_{u\cdot 0}, V) \implies \Big(\bigvee_{0\le j\le i} V = U_j \vee U_i < V\Big)\\
&\wedge x \in U_i.
\end{aligned}$$

Now that our formulae $\varphi_{\mathrm{def}}^{\tau,v}$ are defined for all $v \in \mathrm{Nodes}(\mathrm{Tree}(\tau))$, it suffices, for each

$x \in \lambda$, to define $\varphi_{\text{def}}^{\lambda,x}(y)$ as the formula:

$$\begin{aligned}\exists (C_u)_{u\in\text{Nodes}(\tau)}.\ &\varphi_{\text{decomp}}^{\tau}\big((C_u)_{u\in\text{Nodes}(\tau)}\big)\\ &\wedge \varphi_{\text{piece}}(C_\epsilon, U)\\ &\wedge \varphi_{\text{def}}^{\tau,v}\big((C_u)_{u\in\text{Nodes}(\tau)}, y\big),\end{aligned}$$

with v being the leaf $\iota_\Xi(x)$ of Tree(τ). □

4.6 Conclusion and further work

In this chapter, we were able to show that in the case of finitary linear orders, the only obstacle for uniformisations to be definable in MSO[<] are non-trivial automorphisms. This provides a very clean picture: given a linear order λ, either λ admits a convex subset of the shape $\mu \times \mathbb{Z}$, which does not allow to define even a choice function, or it does not, and in that case all regular relations over λ have regular uniformisations. Moreover, we gave a procedure to construct these regular uniformisations.

The techniques involved in the proof are based mainly on the tools developed in [CCP18] to study the algebraic structure of regular languages of countable words. Yet, our approach slightly differs, in the sense that we focus on a particular linear order, while said article mostly studies languages defined over countable words in general. This highlights that this method of using trees and relying over condensations to find one's way around in the linear orders is rather promising: with only this procedure we were able, in MSO[<], to construct uniformisations, choice functions, well orders, and also to define every position. One might be able to come with the possibility to define new interesting properties of finitary linear orders via these tools.

In our aim to better understand uniformisability over finitary linear orders, one may also wonder which are the regular binary relations that actually admit a regular uniformisation:

Problem 4.32. *Given a finitary linear oder λ and $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^{\lambda}$ a regular binary relation, with $\mathbb{A}$ and $\mathbb{B}$ two alphabets, does R admit a regular uniformisation?*

To achieve this, one should understand how to merge the techniques of [FST20], which analyses the case $\lambda = \mathbb{Z}$, with the above results clarifying the situation under the assumption of "no convex subset of the shape $\mu \times \mathbb{Z}$".

Another question would be about the possibility to define the different objects involved in this chapter in First-Order Logic. An immediate corollary of Proposition 2.1 is that, if λ is an infinite linear order, not all relations $R \subseteq \binom{\mathbb{A}}{\mathbb{B}}^{\lambda}$ definable in $\mathrm{FO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$ admit uniformisations also in $\mathbf{FO}[\binom{\mathbb{A}}{\mathbb{B}}, <]$.

Yet, the author is inclined to believe that, in the context of finitary linear orders, the possibility to define regular choice functions is equivalent to define them in First-Order Logic:

Conjecture 4.33. *Let λ be a finitary linear order. Then the following are equivalent:*

- *λ is rigid,*
- *λ admits a regular choice function,*
- *λ admits an* $\mathrm{FO}[<]$ *choice function.*

Remember that a choice function $f\colon \mathcal{P}(\lambda) \setminus \{\emptyset\} \to \lambda$ is first order (or is an $\mathrm{FO}[<]$ choice function) if it is definable by a formula $\varphi^{\lambda}_{\text{choice}}(X, x)$ that does not contain any second-order quantifier.

So far, the author was only able to prove the conjecture when λ is obtained from singleton sets only and the two operations $(\times\omega)$ and $(\times\omega^{\star})$, meaning when it is of the shape $\xi_0 \times \xi_1 \times \cdots \times \xi_{n-1}$, with each ξ_i being either ω or $\omega^{\star}$. In this case, λ cannot admit any non-trivial automorphism.

Proposition 4.34. *Let λ be a finitary linear order, whose construction does not involve the concatenation operation $+$, nor the η-operation. Then it admits a first-order choice function.*

The proof of this preliminary result is given Appendix C.

Finally, one could ask which of the implications of Theorem 4.1 are preserved when we do not assume the assumption of finitariness: does a countable rigid linear order necessarily admit the regular-uniformisation property? Or maybe it admits a regular choice function? We study these questions in the next chapter.

Chapter 5

The non-finitary case

In the previous chapter, we showed the equivalence between these six different propositions, in the context of finitary linear orders:

- identity is the unique automorphism,
- there is no convex subset isomorphic to $\mu \times \mathbb{Z}$, μ being some non-empty linear order,
- regular-uniformisation property holds,
- regular choice function exists,
- regular well order exists,
- all elements are regularly definable.

Moreover, we proposed algorithms constructing formulae that define those properties, in the case when the first condition holds.

A legitimate question would be whether these equivalences remain true when considering less restricted orders, like countable linear orders not necessarily finitary, or even orders not necessarily linear.

In this chapter, we study these implications in the former case (countable linear orders, not necessarily finitary). The reason we want to keep the assumption of countability is that, with uncountable linear orders, MSO[$<$] does not behave as easily as it does with countable ones. For instance, the MSO[$<$] theory of $\mathbb{R}$, the set of real numbers, is not decidable, in the sense that there does not exist an algorithm that inputs an MSO[$<$] sentence and outputs **YES** if $\mathbb{R} \vDash \varphi$, and **NO** if $\mathbb{R} \nvDash \varphi$ [She75, Theorem 7]—while the MSO[$<$] theory of $\mathbb{Q}$, the set of rational numbers, *is* (once again, see Theorem 6.2. of the same article). Moreover, since every countable linear order is isomorphic to a suborder of $\mathbb{Q}$ (Theorem 1.1), we can deduce from it is decidable whether there exists some countable linear order satisfying a given formula. But this is not the case for non-countable orders.

In Section 5.1, we discuss which of the implications are preserved, while in Section 5.2, we provide counterexamples for some other implications. Figure 5.1 sums up the results of these two sections: the green arrows naturally representing the preserved implications, and the crossed red arrows standing for the unpreserved ones. The latter are labelled by their respective counterexamples.

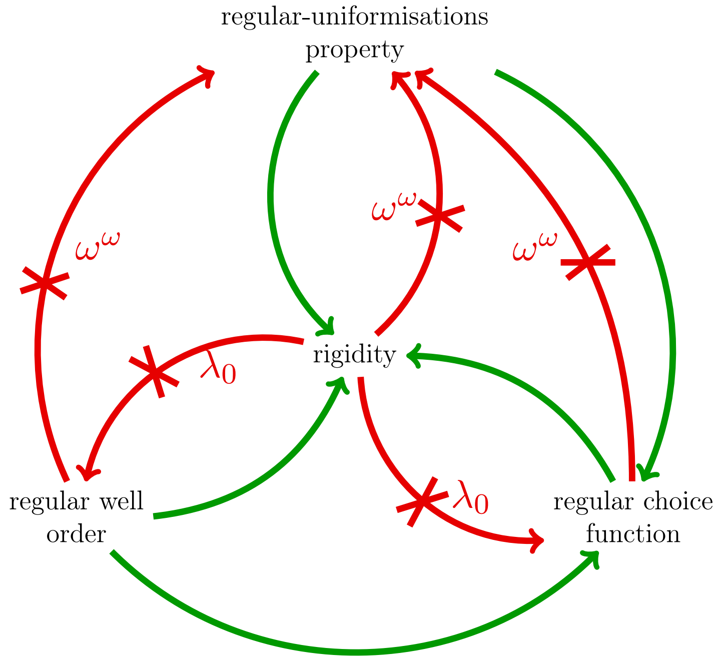


Figure 5.1: The implications and non-implications proven in this chapter, for countable linear orders. The labels ω^ω and λ_0 denote the countable linear orders defined in Section 5.2.

It can be noticed that this figure makes no mention of two of the six conditions we introduced: the non-existence of convex subsets isomorphic to $\mu \times \mathbb{Z}$, and the possibility to define each element regularly. Indeed, it happens that the former condition is equivalent to the non-existence of non-trivial automorphisms, hence there is no need to include it in our graph. Concerning the latter condition, it seemed to the author that, without the assumption of finitariness, this question is not easily related to the other properties, and no sufficiently interesting results were provided.

5.1 Preserved implications

In this short section, we focus on the implications of Theorem 4.1 that remain true without the assumption of λ being finitary. We mostly state anew the implications proven in Section 4.2, since the implications proven there do not require the key assumption of finitariness of a given countable linear order. In fact, they do not require the assumption of countability either, but for the reason we explained at the beginning of this chapter, we do focus on linear orders that are countable.

First, the existence of convex subsets of the shape $\mu \times \mathbb{Z}$ is always equivalent to non-rigidity:

Proposition 5.1. *A countable linear order is non-rigid if and only if it admits a convex subset isomorphic to $\mu \times \mathbb{Z}$, with μ being a non-empty linear order.*

This proposition has already been stated and proven in the previous chapter, as Claim 4.7 (see page 109): a non-trivial automorphism α induces such a convex subset $[x_0, \alpha(x_0)[\times\mathbb{Z}$, with x_0 being any element such that $x_0 < \alpha(x_0)$ (the case $\alpha(x_0) < x_0$ being symmetric), and, reciprocally, a convex subset isomorphic to $\mu \times \mathbb{Z}$ induces a non-trivial automorphism, that maps each copy of μ to its successor.

Concerning the expressive power of MSO[<], we proved also in Section 4.2 that choice was an instance of uniformisation: a linear order λ admits a regular choice function $\varphi^{\lambda}_{\text{choice}}$ if and only if the λ-language of the formula $\varphi_{\text{belongs}} := \forall x.\ \neg\binom{0}{1}(x) \wedge \exists! x.\ \binom{1}{1}(x)$ admits a regular uniformisation. Hence, we could conclude the following implication:

Proposition 5.2. *If a countable linear order satisfies the regular-uniformisation property, then it admits a regular choice function.*

Also, by the very definition of a well order, if λ admits a well order definable in MSO[<], then the choice function that maps each non-empty subset of λ to its least element by this well order is regular. Notice that this is true not only for linear orders, but for any models of any signature, in general:

Fact 5.3. *Let Σ be a signature. If some* MSO[Σ] *formula $\varphi^{\mathcal{M}}_{\text{wo}}(x, y)$ defines a well order over a model $\mathcal{M}$ of Σ, then the formula $\varphi^{\mathcal{M}}_{\text{choice}}(X, x)$ defined as $x \in X \wedge \forall y {\in} X.\ \varphi^{\mathcal{M}}_{\text{wo}}(x, y)$ defines a choice function over $\mathcal{M}$.*

Finally, we recall that the possibility to define all these objects in MSO[<] requires the rigidity assumption:

Proposition 5.4. *If a countable linear order it is non-rigid, then:*
- *it does not admit a regular choice function,*
- *it does not admit a regular well order,*
- *it does not admit the regular-uniformisation property.*

The fact that non-rigidity implies the first two items was already stated in Claim 4.8. The proof uses the fact that satisfiability of formulae is preserved by automorphisms (see Fact 1.5 on page 32). Naturally, we deduce the last implication from Proposition 5.2.

All these preserved implications are depicted in Figure 5.2, which we are willing to complete in the next section.

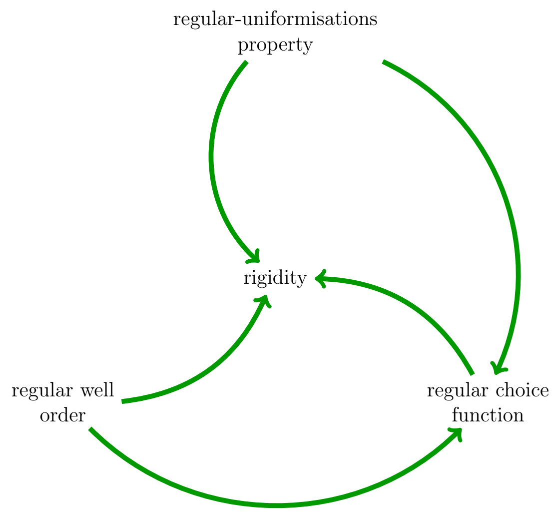


Figure 5.2: A graph of implications between properties of countable linear orders.

5.2 Failing implications

In this second subsection, we highlight some of the implications of Theorem 4.1 which are not preserved without the assumption of finitariness. The first non-implications we state are nothing more than an application of a result of [LS98], which proves that the ordinal ω^ω, among other, does not admit the regular-uniformisation property. Then, we prove a last non-implication by constructing a linear order that is rigid, and yet admits no regular choice function.

5.2.1 The counterexample ω^ω

As stated in this introducing paragraph, the ordinal ω^ω is an example of a countable linear order on which not all regular relations admit regular uniformisations:

Proposition 5.5 ([LS98]). *The linear order ω^ω does not admit the regular-uniformisation property.*

As mentioned on page 28, $(\cdot)^\omega$ shall be understood here as the ordinal-theoretic operation, and this order ω^ω shall not be confused with $\omega \times \omega$ (which does satisfy the regular-uniformisation property, as it is implied by Theorem 4.1 on page 105). Once again, we refer to [Sie58] for a survey of ordinals and that particular operation: ω^ω is defined as the *limit* of the linear orders ω, $\omega^2 = \omega{\times}\omega$, $\omega^3 = \omega{\times}\omega{\times}\omega\dots$ A graphical representation of ω^ω is pictured on Figure 5.3.

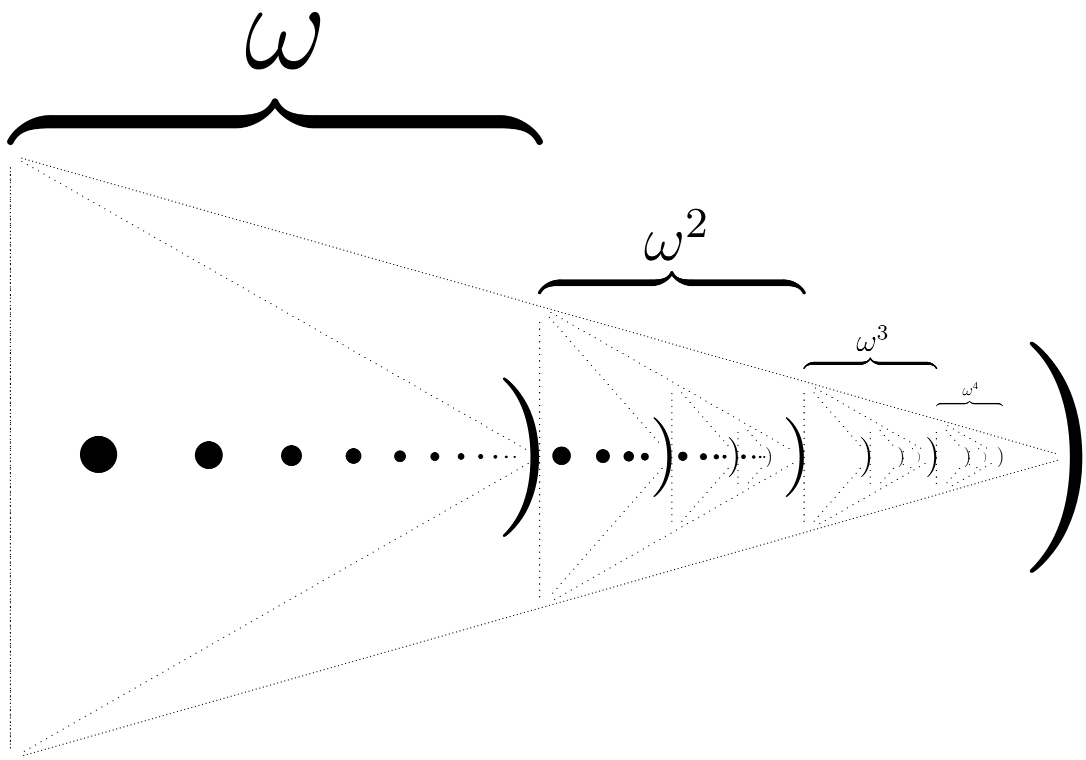


Figure 5.3: A graphical representation of the linear order ω^ω.

Ordinals have the very important property that every well order is isomorphic to a unique ordinal. Hence, ω^ω trivially admits a regular well order, and therefore also a regular choice function, but no non-trivial automorphism (by the implications proven in the previous section). Thus, ω^ω helps up in concluding no less than three non-implications ω^ω:

Proposition 5.6. *The ordinal ω^ω is an example of a countable linear order that:*

- *is rigid,*
- *admits a regular well order,*
- *admits a regular choice function,*
- *and yet does not satisfy the regular-uniformisation property.*

We can complete the previous figure of implications with these negative results, and obtain Figure 5.4.

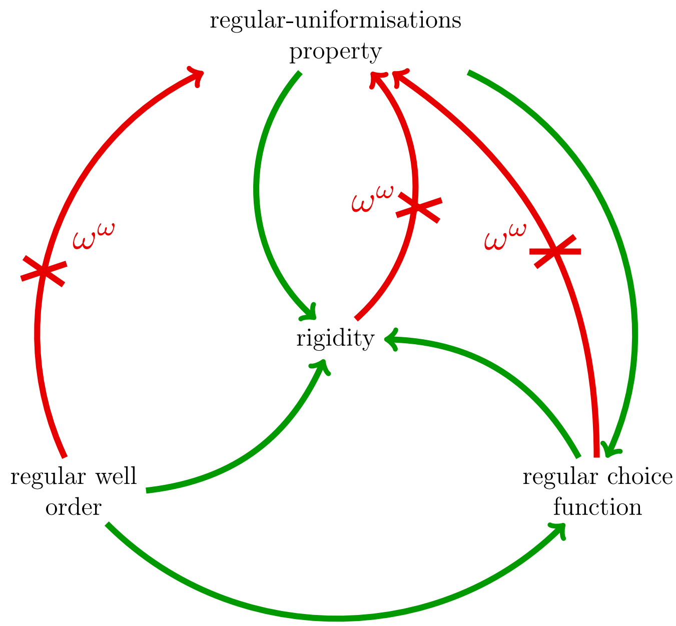


Figure 5.4: Among linear orders, three of these implications fail because of ω^ω.

5.2.2 Rigid, but yet without regular choice function

In this second subsection, we prove that, without the assumption of finitariness, rigidity does not imply any more the existence of regular choice function (and thus nor the existence of regular well order). We prove this negative result by constructing a concrete counterexample.

In order to construct such a linear order, we make use of the fact that the two countable ordinals $\nu_1 := \omega^\omega$ and $\nu_2 := \omega^\omega \times 2$ satisfy exactly the same MSO[<] sentences (*i.e.* for every sentence φ, we have the equivalence $\nu_1 \vDash \varphi$ if and only if $\nu_2 \vDash \varphi$):

Proposition 5.7. *The ordinals ω^ω and $\omega^\omega \times 2$ satisfy the same* MSO[<] *sentences.*

This result is a corollary of Theorem 2 of [Büc65], which proves that two ordinals satisfy the same MSO[<] sentences if and only if they share the same ω-*tail* (a notion which we do not define here).

Notice that, if we did not care about the two ordinals ν_1 and ν_2 to be countable, then we could simply ensure their existence by a simple application of the pigeon-hole principle ("if there are less holes than pigeons, then at least two pigeons must enter the same hole").

Indeed, let ν be some ordinal. We denote by $\mathcal{T}(\nu)$ the set $\{\varphi \in \text{MSO}[<] \mid \nu \vDash \varphi\}$ of the formulae satisfied by it, and we call it its MSO[<] *theory*. The argument is that, considering that formulae in MSO[<] are objects obtained from a finite signature and using a finite amount of finite constructors, there are countably many MSO[<] sentences. This means that the set of sets of sentences has cardinality *continuum*, *i.e.* the cardinality of $\mathcal{P}(\mathbb{N})$, the set of subsets of $\mathbb{N}$ (see [Can74]). It is also a very classic set-theoretical result that the class of ordinals is not a set: it is "too big" to be given any cardinality (see for instance [Jec03, Chapter 2, Section "Ordinal Numbers"]). In particular, there are more ordinals than there are sets of MSO[<] sentences. Thus, by the pigeon-hole principle, there necessarily exist two distinct ordinals ν_1 and ν_2 sharing the same MSO[<] theory: $\mathcal{T}(\nu_1) = \mathcal{T}(\nu_2)$, and $\nu_1 \vDash \varphi$ if and only if $\nu_2 \vDash \varphi$ for all MSO[<] sentences φ.

But since we require the assumption of countability, as we stated at the beginning of the chapter, we consider ν_1 and ν_2 to be ω^ω and $\omega^\omega \times 2$ for the following of the section: in addition to satisfy exactly the same MSO[<] sentences, we need to remember that they are not isomorphic to each other (as distinct ordinals), and infinite.

Now, we define ξ as the linear order $\omega^\star \times \omega$: it consists of an infinite number of copies of $\omega^\star$ (see a graphical representation of it on Figure 5.5).

Unlike ν_1 and ν_2, ξ is not a well order, but, very importantly, it is rigid:

Lemma 5.8. *The linear order ξ is rigid.*

The argument is not of any theoretical interest, but requires some redacting, which we give for the sake of completeness.

Proof. We recall that ξ is, formally, the set of elements of the shape $\langle k, n\rangle$, with n being a natural number, and k being a negative (in the strict sense) integer, and it is ordered by $\langle k_1, n_1\rangle \leq \langle k_2, n_2\rangle$ if either $n_1 < n_2$, or $n_1 = n_2$ and $k_1 \leq k_2$.

Let α be any automorphism of ξ, in order to prove that it is necessarily the identity.

First, notice that if $\alpha\big(\langle k_1, n_1\rangle\big) = \langle k_2, n_2\rangle$, then we have $\alpha\big(\langle k_1{-}1, n_1\rangle\big) = \langle k_2{-}1, n_2\rangle$. Indeed, α conserving the order, we must have $\alpha\big(\langle k_1{-}1, n_1\rangle\big) < \langle k_2, n_2\rangle$, and, if we had $\alpha\big(\langle k_1{-}1, n_1\rangle\big) < \langle k_2{-}1, n_2\rangle$, then $\alpha^{-1}\big(\langle k_2{-}1, n_2\rangle\big)$ would have to be some element x satisfying $\langle k_1{-}1, n_1\rangle < x < \langle k_1, n_1\rangle$, which is impossible.

Hence, to prove that α is the identity, it is sufficient to show that $\alpha\big(\langle -1, n\rangle\big) = \langle -1, n\rangle$ for all natural numbers n. Indeed, if we prove it, then, for any n, applying the previous remark, we also have $\alpha\big(\langle -2, n\rangle\big) = \langle -2, n\rangle$, and then $\alpha\big(\langle -3, n\rangle\big) = \langle -3, n\rangle \ldots$, and $\alpha\big(\langle k, n\rangle\big) = \langle k, n\rangle$ for any negative integer k, by induction.

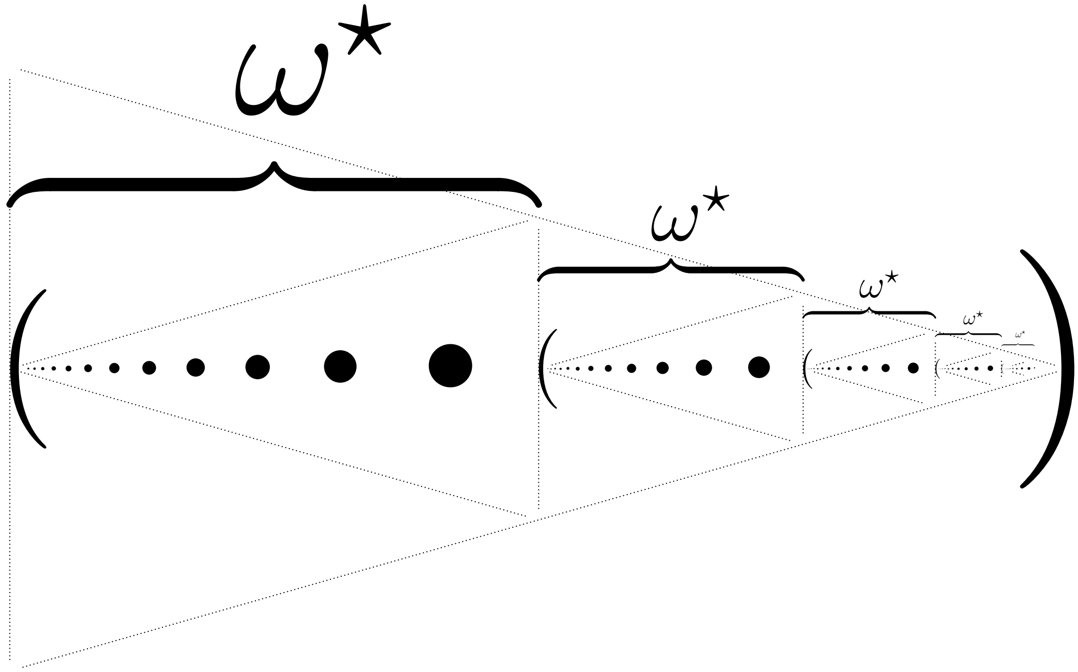


Figure 5.5: A graphical representation of the linear order $\xi = \omega^\star \times \omega$.

Now, by a similar argument as previously, we notice that for any natural number n, the element $\alpha\big(\langle -1, n\rangle\big)$ cannot have any successor in ξ, meaning that it is necessarily the element $\langle -1, m\rangle$, for some $m \in \mathbb{N}$. Indeed, if $\alpha\big(\langle -1, n\rangle\big)$ was of the shape $\langle k_1, m\rangle$, for some $k_1 < -1$ and some $m \in \mathbb{N}$, then $\alpha^{-1}\big(\langle k_1{+}1, m\rangle\big)$ would have to be some element $\langle k_2, p\rangle$, with $n < p$. But then the element $\langle k_2{-}1, p\rangle$, which is between $\langle -1, n\rangle$ and $\langle k_2, p\rangle$, would be mapped to some element x with $\langle k_1, m\rangle < x < \langle k_1{+}1, m\rangle$, which is impossible.

Now that we know that each element of the shape $\langle -1, n\rangle$ has to be mapped to some element $\langle -1, m\rangle$, it remains to be proven that this m is necessarily n itself.

The argument is similar to the one provided to the proof of Example 4.2: let us suppose that there exist some natural numbers n such that $\alpha\big(\langle -1, n\rangle\big) \neq \langle -1, n\rangle$, and let n be the least of them: it is mapped to $\langle -1, m\rangle$, for some $m \neq n$. We cannot have $m < n$, since we would also have $\alpha\big(\langle -1, m\rangle\big) = \langle -1, m\rangle$, which would contradict the injectivity of α. But we cannot have $n < m$ either, because if we did, then $\langle -1, m\rangle$ could not be the image under any element in $\omega^\star \times \omega$, which would contradict the surjectivity of α. Hence, such natural numbers cannot exist, which means that α is the identity function.

We can conclude: $\omega^\star \times \omega$ is rigid. $\square$

In our next step, we can define μ_0 as the linear order $(\nu_1 + \xi) \times \mathbb{Z}$: it is basically a bi-infinite number of copies of ν_1, with a copy of ξ between every two consecutive copies of ν_1. Finally, we define λ_0 as the linear order $(\nu_1{+}\xi){\times}\omega^\star + \nu_2 + (\xi{+}\nu_1){\times}\omega$: we can see it as μ_0, with one of the copies of ν_1 being replaced by a copy of ν_2.

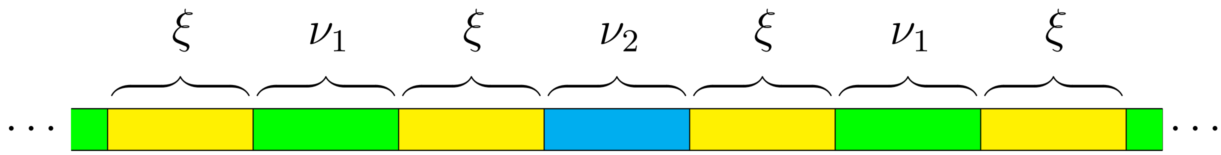


Figure 5.6: A graphical representation of the linear order λ_0.

The idea is that this very copy of ν_2 prevents λ_0 of having any non-trivial automorphisms, but on the other hand, the fact that ν_1 and ν_2 have the same MSO[<] theory implies that λ_0, as μ_0, cannot admit any regular choice function.

Indeed, λ_0 and μ_0 have the same MSO[<] theory:

Claim 5.9. *λ_0 and μ_0 satisfy the same* MSO[<] *sentences.*

Proof. Recall from Subsection 1.2.3 that, if $\mathcal{M}$ and $\mathcal{N}$ are two models over Σ, $\mathcal{M} \cong \mathcal{N}$ means that Duplicator has a winning strategy for the game $\mathcal{G}_d^{\mathrm{MSO}[\Sigma]}(\mathcal{M}, \mathcal{N})$, for every $d \in \mathbb{N}$, and Theorem 1.7 states that this is equivalent to $\mathcal{M}$ and $\mathcal{N}$ satisfying the exact same MSO[Σ] sentences.

The linear order μ_0 is isomorphic to $(\nu_1{+}\xi){\times}\omega^\star + \nu_1 + (\xi{+}\nu_1){\times}\omega$, and we naturally have $(\nu_1{+}\xi){\times}\omega^\star \cong (\nu_1{+}\xi){\times}\omega^\star$. Hence, we deduce $(\nu_1{+}\xi){\times}\omega^\star + \nu_1 \cong (\nu_1{+}\xi){\times}\omega^\star + \nu_2$ by Proposition 1.8 on page 36, since $\nu_1 \cong \nu_2$. Applying once again Proposition 1.8, we obtain $(\nu_1{+}\xi){\times}\omega^\star{+}\nu_1{+}(\xi{+}\nu_1){\times}\omega \cong (\nu_1{+}\xi){\times}\omega^\star{+}\nu_2{+}(\xi{+}\nu_1){\times}\omega$, and therefore $\mu_0 \cong \lambda_0$. $\square$

One can deduce from it that one cannot define in MSO[<] a choice function over λ_0.

Proposition 5.10. *The linear order λ_0 does not admit any regular choice function.*

Proof. Let $\varphi(X, x)$ be any MSO[<] formula with two free variables: one second-order, the other first-order. We show that it cannot define a choice function over λ_0.

For this, we consider the formula ψ_φ, without free variables, defined as

$$\forall X.\ X \neq \emptyset \implies \exists x{\in}X.\ \varphi(X, x) \wedge \forall y.\ \varphi(X, y) \implies x = y.$$

This formula ψ_φ states exactly that $\varphi(X,x)$ defines a choice function: for any linear order λ, $\lambda \vDash \psi_\varphi$ if and only if the binary relation $\big\{\langle X,x\rangle \in \mathcal{P}(\lambda)\times\lambda \mid \lambda \vDash \varphi(X,x)\big\}$ is a choice function over λ.

Since μ_0 is non-rigid, it does not admit any regular choice function, and we have $\mu_0 \nvDash \psi_\varphi$. Thus, $\lambda_0 \nvDash \psi_\varphi$, since $\lambda_0 \cong \mu_0$ (by Claim 5.9), and $\varphi(X,x)$ does not define a choice function over λ_0, which concludes the proof. $\square$

We spend the rest of the section making sure that λ_0 is rigid. Again, this proof is not of any theoretical interest, yet we provide a full proof of it for the sake of completeness. For convenience, we rewrite λ_0 as the set $\xi\times\{0\}\times\mathbb{Z}\sqcup\nu_1\times\{1\}\times\big(\mathbb{Z}\backslash\{0\}\big)\sqcup\nu_2\times\{1\}\times\{0\}$, naturally linearly ordered by:

- $\langle \_,\_,k\rangle < \langle \_,\_,k'\rangle$ for all integers $k<k'$;
- $\langle \_,0,k\rangle < \langle \_,1,k\rangle$ for every integer k;
- $\langle x,0,k\rangle < \langle y,0,k\rangle$ for every integer k and all $x<y$ in ξ;
- $\langle x,1,k\rangle < \langle y,1,k\rangle$ for every integer $k\neq 0$ and all $x<y$ in ν_1;
- $\langle x,1,0\rangle < \langle y,1,0\rangle$ for all $x<y$ in ν_2.

For every $k\in\mathbb{Z}$, we denote by Y_k the subset $\xi\times\{0\}\times\{k\}$ of λ_0. We call it the k-th *ξ-segment.* Similarly, if $k\neq 0$, then the subset $\nu_1\times\{1\}\times\{k\}$ of λ_0, denoted by X_k, is called the k-th *ν_1-segment*, and, finally the subset $\nu_2\times\{1\}\times\{0\}$, denoted by X_0, is called the *ν_2-segment.*

Finally, before going into the details of the proof, we need a notion of cuts and *anti-cuts.* We already define cuts on page 28: a cut of a linear order λ is a non-empty subset C of it that is closed downward, meaning that for all $x<y\in\lambda$, if y belongs to C then x also does. Symmetrically, an *anti-cut* of λ is a non-empty subset C of it that is *closed upward*, meaning that for all $x<y\in\lambda$, if x belongs to C then y also does. Only λ is both a cut and an anti-cut of itself. If λ is ν_1, ν_2, or ξ, we say that a subset C_k of λ_0 is a *cut* (resp. an *anti-cut*) of the λ-segment $\lambda\times\{i\}\times\{k\}$ if C_k is $C\times\{i\}\times\{k\}$, with C being a cut (resp. an anti-cut) of λ.

Remark 5.11. *No cut, nor anti-cut of any ξ-segment of λ_0 is well ordered.*

Proof. It is immediate to see that all cuts and anti-cuts of ξ admit a subset isomorphic to $\omega^\star$, which does not admit any least element. $\square$

Corollary 5.12. *The only convex subsets of λ_0 that are well ordered are finite subsets of ξ-segments, and subsets of ν_1- and ν_2-segments.*

Proof. By the definition of convexity, a convex subset that would intersect both with some ξ-segment and some ν_1- or ν_2-segment would necessarily contain a cut or an anti-cut of this ξ-segment, and would not be well ordered, according to Remark 5.11. Hence, if a convex subset of ξ is a well order, then it is necessarily contained in one of the segments.

If it is contained in some ν_1- or ν_2-segment, then it is indeed well ordered, since ν_1 and ν_2 are. Finally, we easily notice that convex infinite subsets of ξ necessarily contain a subset isomorphic to $\omega^\star$, and therefore cannot be well ordered, which concludes the proof. □

In order to prove that λ_0 is rigid, we prove a first lemma:

Lemma 5.13. *The image of all ν_1- and ν_2-segments under any automorphism of λ_0 are ν_1- and ν_2-segments respectively.*

Proof. We show the claim for a ν_1-segment $X_k = \nu_1 \times \{1\} \times \{k\}$. The proof is similar for the ν_2-segment. The image of X_k under α must be a well order, since isomorphisms preserve well orders. Hence, by Claim 5.12, $\alpha(X_k)$ is necessarily either a finite subset of a ξ-segment, or a subset of a ν_1- or ν_2-segment.

Since X_k is infinite, $\alpha(X_k)$ cannot be finite, and therefore it is necessarily a subset Y of a ν_1- or ν_2-segment Z. We show that we have necessarily $Y = Z$, meaning that it is both a cut and an anti-cut of Z.

Let us show that Y is an anti-cut of Z, by considering the subset $Y' = \{y' \in Z \mid y < y' \text{ for all } y \in Y\}$. If it is non-empty, then $\alpha^{-1}(Y')$, its inverse image under α, necessarily contains a cut of the ξ-segment $\xi \times \{0\} \times \{k+1\}$, and hence is not well ordered. Thus, since Y' is itself a well order, as a subset of a well order, it is necessarily empty, and Y is an anti-cut of Z.

The same way, we prove that Y is a cut of Z, and hence it is Z itself, and we have proven the claim. □

Finally, we can deduce that the identity is the unique automorphism of λ_0, which will conclude our section.

Proposition 5.14. *The linear order λ_0 is rigid.*

Proof. Since ν_1, ν_2, and ξ are all rigid (ν_1 and ν_2 because they are well orders, and ξ because of Claim 5.8), it suffices to prove that for each segment X of λ_0 (ν_1-, ν_2-, or ξ-segment), $\alpha(X)$ is X itself. Indeed, if we prove this, then, for each such segment X, $\alpha\restriction_X$ is an automorphism of X, and hence the identity of X, since X is isomorphic to ν_1, ν_2, or ξ, which do not admit non-trivial automorphisms. Hence, for each $x \in \lambda_0$, by considering X the segment containing x, we have $\alpha(x) = \alpha{\restriction}_X(x) = \mathrm{id}_X(x) = x$: α is the identity.

First, we make sure that $\alpha\big(\nu_2{\times}\{1\}{\times}\{0\}\big)$, the image under α of the ν_2-segment X_0, is the ν_2-segment X_0 itself. Indeed, Lemma 5.13 tells us that it is either a ν_1-segment or the ν_2-segment. But since ν_1 and ν_2 are not isomorphic to each other, it is necessarily X_0.

Now, we know that for all $k \in \mathbb{N}$, the ν_1-segment $X_k = \nu_1{\times}\{1\}{\times}\{k\}$ is mapped to some ν_1-segment $X_{k'}$, with $k' > 0$, by Lemma 5.13 and because α preserves the order. To show that it is necessarily mapped to itself (meaning $k = k'$), we make use of the same argument as in the proofs of Example 4.2 and Lemma 5.8: if there are such ν_1-segments, we consider the least one, which enters into contradiction with the bijectivity of α.

By symmetry, the same arguments apply for the ν_1-segments X_k, with $k < 0$, and we conclude that α maps each ν_1 segment to itself.

Finally, for each $k \in \mathbb{Z}$, the image under α of the ξ-segment $Y_k = \xi{\times}\{0\}{\times}\{k\}$, is the set of elements which are between $\alpha(X_{k-1})$ and $\alpha(X_k)$, meaning between X_{k-1} and X_k, meaning exactly Y_k, and we have shown that the image of each segment of λ_0 is itself. By our remark at the beginning of the proof, we deduce that α is the identity function, and we conclude. □

We have everything to conclude our subsection:

Claim 5.15. *The linear order λ_0 which we constructed is an example of linear order that:*

- *is rigid, and yet*
- *admits no regular choice function,*
- *admits no regular well order.*

Figure 5.1, which we put on page 147, summarises all our results of this chapter.

5.3 Conjectured implications and conclusion

In this chapter, we studied to which extend the equivalences of Theorem 4.1 remain true without the assumption of finitariness, and we were able to give an interesting survey, yet not completed.

Indeed, on summarising Figure 5.1, two (non-)implications are not mentioned: the implication from the regular-uniformisation property to the existence of a regular well order, and the implication from the existence of a regular choice function to the existence of a regular well order.

For now, these two questions have not been answered by the author. Yet, we are tempted to answer them positively. The two of them being a consequence of a stronger conjecture, involving not only linear orders, but general models:

Conjecture 5.16. *Let Σ be a signature. For each* $\mathrm{MSO}[\Sigma]$ *formula $\varphi_{\mathrm{choice}}(X, x)$, there exists a formula $\varphi_{\mathrm{wo}}(x, y)$, also in* $\mathrm{MSO}[\Sigma]$*, such that for every model $\mathcal{M}$ of Σ, if $\varphi_{\mathrm{choice}}(X, x)$ defines a choice function over $\mathcal{M}$, then $\varphi_{\mathrm{wo}}(x, y)$ defines a well order over $\mathcal{M}$.*

Hence, would any linear order λ admit a regular choice function, defined by an $\mathrm{MSO}[<]$ formula $\varphi^{\lambda}_{\mathrm{choice}}(X, x)$, it would also admit a regular well order, defined by a formula $\varphi^{\lambda}_{\mathrm{wo}}(x, y)$, depending only on $\varphi^{\lambda}_{\mathrm{choice}}$ (and not on λ).

Even though Conjecture 5.16 is unproved so far, it is worth stating a related lemma, which could help us on the matter. Let Σ be a signature, and $\psi(x, y)$ be an $\mathrm{MSO}[\Sigma]$ formula over Σ, with two free first-order variables. We say that ψ *defines a choice function* over a model $\mathcal{M}$ of Σ if $\mathcal{M} \vDash \forall X{\neq}\emptyset.\ \exists! x{\in}X.\ \forall y{\in}X.\ \psi(x, y)$, meaning that for each non-empty subset X of $\mathcal{M}$, there exists a unique $x \in X$ such that $\mathcal{M} \vDash \psi(x, y)$ for all $y \in X$. The following lemma tells that such a formula defines a well order as well:

Lemma 5.17. *Let Σ be a signature, let $\mathcal{M}$ be a model of Σ, and let $\psi(x, y)$ be an* $\mathrm{MSO}[\Sigma]$ *formula, with two free first-order variables x and y. Then, if $\psi(x, y)$ defines a choice function over $\mathcal{M}$, then it also defines a well order over $\mathcal{M}$.*

Proof. Let $R^{\mathcal{M}}$ be the binary relation over $\mathcal{M}$ defined by ψ: $R^{\mathcal{M}} := \big\{\langle x, y\rangle \mid \mathcal{M} \vDash \psi(x, y)\big\}$, and let us suppose that $\psi(x, y)$ defines a choice function over $\mathcal{M}$: for every non-empty subset X of $\mathcal{M}$, there exists a unique $x \in X$ such that $xR^{\mathcal{M}}y$ for all $y \in X$.

It is sufficient for us to prove that $R^{\mathcal{M}}$ is an order. Indeed, if it is, then the assumption of ψ ensures the existence of a least element with respect to $R^{\mathcal{M}}$ for every non-empty subset, which is exactly the definition of a well order.

- First, reflexivity. Let $x \in \mathcal{M}$: the unique element $x' \in \{x\}$ such that $x'R^{\mathcal{M}}y$ for all $y \in \{x\}$ is necessarily x, and therefore $xR^{\mathcal{M}}x$.
- Second, antisymmetricity. Let $x, y \in \mathcal{M}$, and let us suppose that $xR^{\mathcal{M}}y$ and $yR^{\mathcal{M}}x$. By reflexivity, we also have $xR^{\mathcal{M}}x$ and $yR^{\mathcal{M}}y$. Therefore, x and y are two elements x' of $\{x, y\}$ such that $x'R^{\mathcal{M}}y'$ for all $y' \in \{x, y\}$. By unicity of this element x', we necessarily have $x = y$.
- Finally, transitivity. Let $x, y, z \in \mathcal{M}$, and let us suppose that $xR^{\mathcal{M}}y$ and $yR^{\mathcal{M}}z$. If $x = y$ or $y = z$, then we clearly have $xR^{\mathcal{M}}z$ by assumption. If $x = z$, then we have it by reflexivity. Hence, let us assume that these three elements are pairwise distinct. By

antisymmetricity, this means that we do not have $yR^{\mathcal{M}}x$, nor $zR^{\mathcal{M}}y$. Hence, neither y nor z can be elements $x' \in \{x, y, z\}$ such that $x'R^{\mathcal{M}}y'$ for every $y' \in \{x, y, z\}$. This means that this x' is necessarily x, and therefore $xR^{\mathcal{M}}z$.

We showed that $R^{\mathcal{M}}$ is an order over $\mathcal{M}$, and this concludes our proof. $\square$

Hence, if for every MSO[Σ] formula $\varphi_{\text{choice}}(X, x)$, we managed to show that there exists an MSO[Σ] formula $\varphi_{\text{wo}}(x, y)$ such that the formula:

$$\forall X{\neq}\emptyset.\ \exists x{\in}X.\ \varphi_{\text{choice}}(X, x) \wedge \forall x'{\in}X'.\ \varphi_{\text{choice}}(X, x') \implies x = x'$$

is logically equivalent to the formula:

$$\forall X{\neq}\emptyset.\ \exists x{\in}X.\ \forall y{\in}X.\ \varphi_{\text{wo}}(x, y) \wedge \forall x'{\in}X.\ \big(\forall y{\in}X.\ \varphi_{\text{wo}}(x', y)\big) \implies x = x',$$

then, under the condition that φ_{choice} defines a choice function over a model, said φ_{wo} would define a well order over it, and we would be able to conclude Conjecture 5.16.

We shall put as a remark that there is *a priori* no reason to deduce such an eventual formula $\varphi_{\text{wo}}(x, y)$ from the semantics of $\varphi_{\text{choice}}(X, x)$, since said semantics are often difficult to comprehend. For instance, a regular choice over ω would be (among other) the function that maps every non-empty subset to either its greatest element if it has one (meaning it is finite), or its least element if not. It is indeed definable by a formula $\varphi^{\omega}_{\text{choice}}(X, x)$. The well order we would like to obtain from it is the order of ω, *i.e.* $\varphi_{\text{wo}}(x, y)$ should be equivalent to $x \leq y$, but, in a sense, $\varphi^{\omega}_{\text{choice}}$ does exactly the opposite: for every two elements x and y, we have $\omega \vDash \varphi^{\omega}_{\text{choice}}(\{x, y\}, x)$ if and only if $y \leq x$. This highlights that, in the general case, the obtention of $\varphi_{\text{wo}}(x, y)$ should be syntactical: it should rely on the formula $\varphi_{\text{choice}}(X, x)$ itself, rather than on its semantic interpretation.

So far, the author was not able to come with an algorithm that inputs an MSO[Σ] formula $\varphi_{\text{choice}}(X, x)$ and outputs such a formula $\varphi_{\text{wo}}(x, y)$, but yet believes that this is a good direction to consider, in order to solve Conjecture 5.16.

Naturally, since the regular-uniformisation property implies the existence of a regular choice function, proving Conjecture 5.16 would also prove the last missing implication of Figure 5.1.

Conjecture 5.18. *If a (countable) linear order satisfies the regular-uniformisation property, then it admits a regular well order.*

Concerning these conjectures, it is worth citing the case of trees. It was indeed shown in [GS83, CLNW10] that the infinite binary tree does not admit any well order, nor choice functions that can be defined in Monadic Second-Order Logic.

Conclusions

The main motivating question of this thesis is: under which assumption is it possible to uniformise relations in a given formalism? We focused on this question in the realm of formal languages, both of finite words and of words of countable domains.

Overall, the thesis can be seen as a search for generalisations of Theorems 1.48 and 1.49, stating that the formalism of Monadic Second-Order Logic does allow the construction of uniformisations for regular languages, of respectively finite and ω-words. Many of the provided arguments and constructions rely on an interplay between the semantic assumption of uniformisability and algebraic tools used to describe the expressive power of the considered formalisms.

In Chapter 2, we pointed out that, already over finite words, most of natural fragments of First-Order Logic are too weak to uniformise not only their own relations, but also relations of even weaker fragments. In order to better understand the reasons for uniformisability in these formalisms, we provided an algorithm that inputs a regular relation and outputs (if there exists any) a uniformisation of it in $\mathbf{FO}[\,]$, the fragment of First-Order Logic that can only test letters. It would be interesting to see if the algorithm could be adapted to stronger fragments, like $\mathbf{FO}^2[<]$, $\mathbf{FO}[\mathsf{s}]$, $\mathbf{FO}[<]$...

In Chapter 3, we were able to show that, among non-empty varieties of languages, only **MSO** is actually able to uniformise all its relations. The author believes that this result is rather theoretically important, as it reveals strong connections between the question of uniformisation and the expressive power of Monadic Second-Order Logic.

Finally, in Chapters 4 and 5, we go beyond the domain of finite words and study the question of uniformisation for countable words. The main theorem of Chapter 4, Theorem 4.1, which can be seen as a generalisation of Theorem 1.49, highlights that, in the case of finitary linear orders, the only obstacle to the possibility to define uniformisations in MSO is the existence of non-trivial automorphisms, that "shift" the linear order. The theorem also highlights the equivalence between rigidity (the fact of not admitting such non-trivial automorphisms), regular uniformisations, regular choice functions, regular well orders, and regular definitions of the positions. In Chapter 5, we provided some counterexamples showing that not all these questions remain equivalent without the finitariness assumption. In particular, the assumption of rigidity remains crucial, but not sufficient in itself.

To summarise, the author hopes that this thesis provides a rather complete landscape on uniformisations and choice functions in the field of formal languages. It features a coherent and homogeneous study over a fundamental question, declined in many different ways, and brings its stone to the already long-existing and rich building that makes the connection between expressive powers of formalisms and algebraic structures. Furthermore, it raises new interesting questions in various domains such as model theory or theory of linear orders.

Bibliography

[Ban32] Stefan Banach. *Théorie des opérations linéaires*. Monografie Matematyczne. 1932.

[Bar13] Kevin Barnum. The axiom of choice and its implications. `https://math.uchicago.edu/~may/REU2014/REUPapers/Barnum.pdf`, 2013.

[BL67] Julius Richard Büchi and Lawrence H. Landweber. Solving sequential conditions by finite state strategies. *Department of Computer Science Technical Reports*, 138(88), 1967.

[BP91] Danièle Beauquier and Jean-Éric Pin. Languages and scanners. *Theoretical Computer Science*, 84(1):3–21, 1991.

[BT24] Stefan Banach and Alfred Tarski. Sur la décomposition des ensembles de points en parties respectivement congruentes. *Fundamenta Mathematicae*, 6(1):244–277, 1924.

[Büc60] Julius Richard Büchi. Weak second-order arithmetic and finite automata. *Mathematical Logic Quarterly*, 6(1-6):66–92, 1960.

[Büc65] Julius Richard Büchi. Decision methods in the theory of ordinals. *Bulletin of the American Mathematical Society*, 71(5):767 – 770, 1965.

[Can74] Georg Cantor. Über eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen. *Journal für die reine und angewandte Mathematik*, (77):258 – 262, 1874.

[CCP18] Olivier Carton, Thomas Colcombet, and Gabriele Puppis. An algebraic approach to MSO-definability on countable linear orderings. *J. Symb. Log.*, 83(3):1147–1189, 2018.

[CG99] Christian Choffrut and Serge Grigorieff. *Uniformization of Rational Relations*, pages 59–71. Springer Berlin Heidelberg, 1999.

[CL07] Arnaud Carayol and Christof Löding. MSO on the infinite binary tree: Choice and order. In *CSL*, pages 161–176, 2007.

[CLNW10] Arnaud Carayol, Christof Löding, Damian Niwiński, and Igor Walukiewicz. Choice functions and well-orderings over the infinite binary tree. *Central European Journal of Mathematics*, 8:662–682, 2010.

[Coh63] Paul J. Cohen. The independence of the axiom of choice. *Proceedings of the U.S. National Academy of Sciences*, page 1143–48, 1963.

[CPP08] Olivier Carton, Dominique Perrin, and Jean-Éric Pin. Automata and semigroups recognizing infinite words. In *Logic and Automata: History and Perspectives [in Honor of Wolfgang Thomas]*, volume 2 of *Texts in Logic and Games*, pages 133–168. Amsterdam University Press, 2008.

[CPS06] Laura Chaubard, Jean-Éric Pin, and Howard Straubing. First order formulas with modular predicates. In *21th IEEE Symposium on Logic in Computer Science (LICS 2006)*, pages 211–220. IEEE Computer Society, 2006.

[DFKL20] Vrunda Dave, Emmanuel Filiot, Shankara Narayanan Krishna, and Nathan Lhote. Synthesis of Computable Regular Functions of Infinite Words. In *31st International Conference on Concurrency Theory (CONCUR 2020)*, volume 171 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 43:1–43:17. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020.

[DGK08] Volker Diekert, Paul Gastin, and Manfred Kufleitner. A survey on small fragments of first-order logic over finite words. *Int. J. Found. Comput. Sci.*, 19(3):513–548, 2008.

[EF95] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite model theory. Perspectives in mathematical logic.* Springer, 1995.

[Ehr61] Andrzej Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49(2):129–141, 1961.

[Eil74] Samuel Eilenberg. *Automata, languages, and machines.* Pure and Applied Mathematics. Elsevier Science, 1974.

[Elg61] Calvin C. Elgot. Decision problems of finite automata design and related arithmetics. *Transactions of the American Mathematical Society*, 98(1):21–51, 1961.

[Fra55] Roland Fraïssé. *Sur quelques classifications des systèmes de relations*. Publications Scientifiques de l'Universite d'Alger. Imprimerie Durand, 1955.

[FST20] Grzegorz Fabiański, Michał Skrzypczak, and Szymon Toruńczyk. Uniformisations of regular relations over bi-infinite words. In *LICS '20: 35th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 384–396. ACM, 2020.

[Göd38] Kurt Gödel. The consistency of the axiom of choice and of the generalized continuum-hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 24(12):556—557, December 1938.

[GS83] Yuri Gurevich and Saharon Shelah. Rabin's uniformization problem. *J. Symb. Log.*, 48(4):1105–1119, 1983.

[Hau08] Felix Hausdorff. Grundzüge einer theorie der geordneten mengen. *Mathematische Annalen*, 65:435–505, 1908.

[HMU06] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation (3rd Edition)*. Addison-Wesley Longman Publishing Co., Inc., 2006.

[Imm99] Neil Immerman. *Descriptive complexity*. Springer, 1999.

[Jec03] Thomas Jech. *Set Theory: The Third Millennium Edition*. Springer, 2003.

[Kam68] Hans Kamp. *Tense Logic and the Theory of Linear Order*. PhD thesis, Ucla, 1968.

[Kec95] Alexander S. Kechris. *Classical Descriptive Set Theory*. Graduate texts in mathematics. Springer-Verlag, 1995.

[Kob69] Kojiro Kobayashi. Classification of formal languages by functional binary transductions. *Information and Control*, 15(1):95–109, 1969.

[Kur22] Kazimierz Kuratowski. Une méthode d'élimination des nombres transfinis des raisonnements mathématiques. *Fundamenta Mathematicae*, 3(1):76–108, 1922.

[Leb02] Henri-Léon Lebesgue. *Intégrale, longueur, aire*. Bernandon de C. Rebeschini, 1902.

[LL66] Hans Läuchli and Jon N. Leonard. On the elementary theory of linear order. *Fundamenta mathematicae*, 59:109–116, 1966.

[LMS19] Nathan Lhote, Vincent Michielini, and Michał Skrzypczak. Uniformisation gives the full strength of regular languages. In *44th International Symposium on Mathematical Foundations of Computer Science, MFCS 2019*, volume 138 of *LIPIcs*, pages 61:1–61:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[Lot97] M. Lothaire. *Combinatorics on Words*. Cambridge Mathematical Library. Cambridge University Press, 2md edition, 1997.

[LS98] Shmuel Lifsches and Saharon Shelah. Uniformization and Skolem functions in the class of trees. *J. Symb. Log.*, 63(1):103–127, 1998.

[Mic18] Vincent Michielini. Uniformization problem for variants of first order logic over finite words. In *Developments in Language Theory - 22nd International Conference, DLT 2018*, volume 11088 of *Lecture Notes in Computer Science*, pages 516–528. Springer, 2018.

[Mos80] Yiannis N. Moschovakis. *Descriptive set theory*, volume 100 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., 1980.

[MS20] Vincent Michielini and Michał Skrzypczak. Regular choice functions and uniformisations for countable domains. In *45th International Symposium on Mathematical Foundations of Computer Science, MFCS 2020*, volume 170 of *LIPIcs*, pages 69:1–69:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[Ner58] Anil Nerode. Linear automaton transformations. *Proceedings of the American Mathematical Society*, 9(4):541–544, 1958.

[Pin12] Jean-Éric Pin. Equational descriptions of languages. *Int. J. Found. Comput. Sci.*, 23(6):1227–1240, 2012.

[Pin20] Jean-Éric Pin. Mathematical foundations of automata theory. `https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf`, 2020.

[PS16] Thomas Place and Luc Segoufin. Decidable characterization of FO2($<, +1$) and locality of DA. *CoRR*, abs/1606.03217, 2016.

[PW97] Jean-Éric Pin and Pascal Weil. Polynominal closure and unambiguous product. *Theory Comput. Syst.*, 30(4):383–422, 1997.

[Rab69] Michael Oser Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.

[Rab70] Michael Oser Rabin. Weakly definable relations and special automata. In *Proceedings of the Symposium on Mathematical Logic and Foundations of Set Theory*, pages 1–23. North-Holland, 1970.

[Rab07] Alexander Rabinovich. On decidability of monadic logic of order over the naturals extended by monadic predicates. *Information and Computation*, 205(6):870–889, 2007.

[Rau09] Wolfgang Rautenberg. *A Concise Introduction to Mathematical Logic*. Springer Publishing Company, Incorporated, 3rd edition, 2009.

[Rei82] Jan Reiterman. The birkhoff theorem for finite algebras. *algebra universalis*, 14:1–10, 1982.

[Ros81] Joseph G. Rosenstein. *Linear orderings*. Academic Press New York, 1981.

[RS59] Michael Oser Rabin and Dana Scott. Finite automata and their decision problems. *IBM Journal of Research and Development*, 3(2):114–125, April 1959.

[RS08] Alexander Rabinovich and Amit Shomrat. Selection in the monadic theory of a countable ordinal. *Journal of Symbolic Logic*, 73(3):783–816, 2008.

[Rus19] Bertrand Russell. *Introduction to Mathematical Philosophy*. George Allen & Unwin, Ltd., 1919.

[Sch65] Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8(2):190–194, 1965.

[Sha91] Stewart Shapiro. *Foundations Without Foundationalism: A Case for Second-Order Logic*. Oxford University Press, 1991.

[She75] Saharon Shelah. The monadic theory of order. *The Annals of Mathematics*, 102(3):379–419, 1975.

[Sie58] Wacław Sierpiński. *Cardinal and Ordinal Numbers*. Państwowe Wydawnictwo Naukowe, 1958.

[Sim84] Imre Simon. Word Ramsey theorems. In *Graph Theory and Combinatorics*, pages 283–291. London Academic Press, 1984.

[Tar33] Alfred Tarski. *Pojęcie prawdy w językach nauk dedukcyjnych*. Prace Towarzystwa Naukowego Warszawskiego Wydział III – Nauk matematyczno-fizycznych. Nakł. Tow. Naukowego Warszawskiego, 1933.

[Tra62] Boris A. Trakhtenbrot. Finite automata and the monadic predicate calculus. *Siberian Mathematical Journal*, 3(1):103–131, 1962.

[TW98] Denis Thérien and Thomas Wilke. Over Words, Two Variables Are as Powerful as One Quantifier Alternation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 234–240. ACM, 1998.

[Tyc30] Andrey Tychonoff. Über die topologische erweiterung von räumen. *Mathematische Annalen*, 102:544–561, 1930.

[Vit05] Giuseppe Vitali. *Sul problema della misura dei Gruppi di punti di una retta*. Tip. Gamberini e Parmeggiani, 1905.

[Whi78] Thomas A. Whitelaw. *An Introduction to Abstract Algebra*. Blackie, 1978.

[WI07] Philipp Weis and Neil Immerman. Structure theorem and strict alternation hierarchy for FO2 on words. In *Computer Science Logic*, pages 343–357. Springer Berlin Heidelberg, 2007.

[Zer04] Ernst Zermelo. Beweis, daß jede menge wohlgeordnet werden kann. (aus einem an herrn hilbert gerichteten briefe). *Mathematische Annalen*, 59:514–516, 1904.

[Zor35] Max Zorn. A remark on method in transfinite algebra. *Bulletin of the American Mathematical Society*, 41(10):667–670, 10 1935.

Appendices

A Fragments of First-Order Logic being varieties of languages

In this appendix, we give a proof of Proposition 1.25, on page 47, stating that the formalisms pictured on Figure 1.1, on page 40, are varieties of languages. Recall that a class of languages of finite words is a variety of languages if it is closed under Boolean combinations, quotients, and preimages under homomorphisms (see the definition on page 45).

Proposition 1.25. *The classes of languages of finite words pictured on Figure 1.1, on page 40, are all varieties of languages.*

Proposition 1.23 on page 46, tells us that the class $\mathbf{MSO}[<]$ of regular languages is a variety of languages. In Subsection A.1, we deduce the fact that $\mathbf{FO}[<]$ is a variety, as well as a few of its fragments, from Reiterman's theorem. Finally, we take care of fragments $\mathbf{FO}^k[\,]$ and $\mathbf{FO}^k[\mathsf{s}]$ in the next two subsections. An intersection of varieties naturally being a variety, we will be able to deduce that also the class $\mathbf{FO}^2[<] \cap \mathbf{FO}^2[\mathsf{s}]$ is one.

A.1 Using Reiterman's theorem

To show that $\mathbf{FO}[<]$, First-Order Logic over finite words, is a variety of languages, we rely on Reiterman's theorem [Rei82], which proves that classes of finite languages that can be described via algebraic equations over finite semigroups are varieties of languages. As it happens that languages definable in FO[<] are the ones whose syntactic finite semigroup is aperiodic, *i.e.* satisfies the equation $s^\sharp * s = s^\sharp$, as we stated it in Theorem 1.19 on page 44, we can deduce the result.

Similarly, the classes $\mathbf{FO}[\mathsf{s}]$ and $\mathbf{FO}^2[<,\mathsf{s}]$ are varieties of languages, as Theorems 1.20 and 1.22 describe them via algebraic equations. We can also cite [TW98, Theorem 6], which provides an algebraic characterisation for $\mathbf{FO}^2[<]$, and conclude that this class is a variety of languages. Finally, on page 75, we proved in Theorem 2.22 that $\mathbf{FO}[\,]$ corresponds to aperiodic and commutative finite semigroups, *i.e.* finite semigroups that satisfy the algebraic equations $s^\sharp * s = s^\sharp$ and $s * t = t * s$:

Proposition A.1. *The classes* $\mathbf{FO}[<]$, $\mathbf{FO}[\mathsf{s}]$, $\mathbf{FO}^2[<,\mathsf{s}]$, $\mathbf{FO}^2[<]$, *and* $\mathbf{FO}[\,]$ *are varieties of languages.*

A.2 Classes $\mathbf{FO}^k[\,]$ are varieties of languages

In this subsection, we show that for each natural number k, the class $\mathbf{FO}^k[\,]$ of languages that can be defined with at most k variables and only letter tests is a variety of languages. For this purpose, we will mostly rely on the semantic characterisation of $\mathbf{FO}^k[\,]$, which we proved in Section 2.2. In order to keep the notations consistent with said section, we will rather say that N is the number of allowed variables, and we will denote the class by $\mathbf{FO}^{k=N}[\,]$.

First, it is immediate that $\mathbf{FO}^{k=N}[\,]$ is closed under Boolean combinations, since the constructors $\neg$ and $\vee$ are part of the grammar. It remains to be proven that it is also closed under quotients and preimages under homomorphisms.

Theorem 2.17, which we proved on page 71, states that a language L over an alphabet $\mathbb{A}$ is in $\mathbf{FO}^{k=N}[\mathbb{A}]$ if and only if it is in $\mathbf{M}_N[\mathbb{A}]$, meaning that it can be written as a union $\bigsqcup_{m\in\Lambda} L_m$, with Λ being a set of $\langle\mathbb{A},N\rangle$-maps. We refer to page 70 for a definition of an $\langle\mathbb{A},N\rangle$-map m and of its corresponding language L_m.

We can define L_m as the intersection of the languages $L_{m,a}$, where, for each letter a in $\mathbb{A}$, $L_{m,a}$ is the language of words over $\mathbb{A}$ whose number of occurrences of a match with $m(a)$, in the sense that:

- if $m(a) \in N$, then $L_{m,a}$ is the language $L_a^{=m(a)} := \{w \in \mathbb{A}^+ \mid |w|_a = m(a)\}$;
- if $m(a)$ is $\trianglerighteq_N$, then $L_{m,a}$ is the language $L_a^{\geq N} := \{w \in \mathbb{A}^+ \mid |w|_a \geq N\}$.

We take a brief moment to notice that, if $(L_i)_{i\in I}$ is any family of languages over $\mathbb{A}$, and if u is a finite word over $\mathbb{A}$, then $u^{-1}\cdot\big(\bigcap_{i\in I} L_i\big)$ and $\big(\bigcap_{i\in I} L_i\big)\cdot u^{-1}$, the quotient languages of the intersection $\bigcap_{i\in I} L_i$, are equal to the intersections of the quotient languages $\bigcap_{i\in I}\big(u^{-1}\cdot L_i\big)$ and $\bigcap_{i\in I}(L_i \cdot u^{-1})$ respectively. Hence, since $\mathbf{M}_N[\mathbb{A}]$ is closed under Boolean combinations, it is enough to prove, in order to conclude that it is also closed under quotients, that the languages $u^{-1}{\cdot}L_{m,a}$ and $L_{m,a}{\cdot}u^{-1}$ can be obtained as unions of $\langle\mathbb{A},N\rangle$-maps, for every $\langle\mathbb{A},N\rangle$-map m and every letter $a \in \mathbb{A}$. Thus, we consider an $\langle\mathbb{A},N\rangle$-map m, a letter $a \in \mathbb{A}$, and a word $u \in \mathbb{A}^+$, and we write down the argument for $u^{-1}{\cdot}L_{m,a}$ (the argument for $L_{m,a}{\cdot}u^{-1}$ being identical).

If $m(a) \in N$, then $L_{m,a}$ is the language $L_a^{=m(a)}$ of words over $\mathbb{A}$ that admit exactly $m(a)$ occurrences of a, and therefore $u^{-1}{\cdot}L_{m,a}$ is either the empty language $\emptyset$ if $m(a) < |u|_a$, or the language $L_a^{=m(a)-|u|_a}$ if $m(a) \geq |u|_a$. In both cases, it is expressible via $\langle \mathbb{A}, N \rangle$-maps.

If $m(a)$ is $\trianglerighteq_N$, then $L_{m,a}$ is the language $L_a^{\geq N}$ of words over $\mathbb{A}$ that admit at least N occurrences of a, and hence $u^{-1}{\cdot}L_{m,a}$ is either the full language $\mathbb{A}^+$ if $N \leq |u|_a$, or the language $L_a^{\geq N-|u|_a}$ if $N > |u|_a$. In the first case, it is naturally expressible via $\langle \mathbb{A}, N \rangle$-maps (it is the union of all these maps). In the second case, it is also: indeed, it can be written as the union of the maps m' that satisfy $m'(a) \in \{N{-}|u|_a, N{-}|u|_a{+}1, \ldots, N{-}1, \trianglerighteq_N\}$. We can conclude:

Lemma A.2. *For every natural number N, the class $\mathbf{FO}^{k=N}[\;]$ is closed under quotients.*

Now, it remains to be proven that $\mathbf{FO}^{k=N}[\;]$ is closed under preimages under homomorphisms. Here, again, we notice that, if $(L_i)_{i \in I}$ is a family of languages over an alphabet $\mathbb{B}$, and h a homomorphism from $\mathbb{A}^+$ to $\mathbb{B}^+$, then the languages $h^{-1}\big(\bigcap_{i \in I} L_i\big)$ and $h^{-1}\big(\bigcup_{i \in I} L_i\big)$ are equal to the languages $\bigcap_{i \in I} h^{-1}(L_i)$ and $\bigcup_{i \in I} h^{-1}(L_i)$ respectively. Hence, in the same way, in order to show this closure of $\mathbf{FO}^{k=N}[\;]$, it is enough to prove that if N is a natural number, m is an $\langle \mathbb{B}, N \rangle$-map, and b is a letter of $\mathbb{B}$, then the language $h^{-1}(L_{m,b})$ is in $\mathbf{M}_N[\mathbb{A}]$. Let us prove it.

In a first step, let us assume that $m(b)$ is a natural number smaller than N. Then $L_{m,b}$ is $L_b^{=m(b)}$, the language of words over $\mathbb{B}$ that admit exactly $m(b)$ occurrences of b. A first thing to notice it that if w is a word in $h^{-1}(L_{m,b})$, then all the letters a that have an occurrence in w necessarily satisfy $|h(a)|_b \leq |h(w)|_b = m(b) \leq N{-}1$. Furthermore, such a letter a satisfies either $|h(a)|_b = 0$, or $|w|_a \leq N{-}1$, because $|w|_a \times |h(a)|_b \leq |h(w)|_b = m(b) \leq N{-}1$.

Knowing this, for each $a \in \mathbb{A}$, the natural number $|h(a)|_b$ appears like a key number, and we can part $\mathbb{A}$ into three subsets: the subset $\mathbb{A}_0^b(h)$ of letters a that satisfy $|h(a)|_b = 0$, the subset $\mathbb{A}_{\text{btwn}}^b(h)$ of letters a that satisfy $1 \leq |h(a)|_b < N$, and the subset $\mathbb{A}_{\trianglerighteq_N}^b(h)$ of letters a that satisfy $N \leq |h(a)|_b$.

From the commentaries above, we can characterise the words $w \in \mathbb{A}^+$ that belong to the language $h^{-1}(L_{m,b})$. Such a word w can be described via a function $f_w \colon \mathbb{A}_{\text{btwn}}^b(h) \to \{0, \ldots, N{-}1\}$ that satisfies $\Sigma_{a \in \mathbb{A}_{\text{btwn}}^b(h)} f_w(a) \times |h(a)|_b = m(b)$: w admits any number of occurrences of letters in $\mathbb{A}_0^b(h)$, no occurrences of letters in $\mathbb{A}_{\trianglerighteq_N}^b(h)$, and, for each letter a in $\mathbb{A}_{\text{btwn}}^b(h)$, it admits exactly $f_w(a)$ occurrences of a. This characterisation is expressible via a Boolean combination of $\langle \mathbb{A}, N \rangle$-maps, and we have our result.

The case when $m(a)$ is the symbol $\trianglerighteq_N$ is similar. Here, $f_w \colon \mathbb{A}_{\text{btwn}}^b(h) \to \{0, \ldots, N{-}1\}$ is a function that satisfies the inequality $\Sigma_{a \in \mathbb{A}_{\text{btwn}}^b(h)} f_w(a) \times |h(a)|_b \geq N$, and any word w

in $h^{-1}(L_{m,b})$ can admit either any positive number of occurrences of any letter $a \in \mathbb{A}^b_{\trianglerighteq N}(h)$, or at least N occurrences of any letter in $\mathbb{A}^b_{\text{btwn}}(h)$, or finally exactly $f_w(a)$ occurrences of each letter a in $\mathbb{A}^b_{\trianglerighteq N}(h)$. Naturally, in addition, it can admit any number of occurrences of letters in $\mathbb{A}^b_0(h)$. These conditions are also expressible via Boolean combinations of $\langle \mathbb{A}, N\rangle$-maps, and we can conclude:

Lemma A.3. *For every natural number N, the class $\mathbf{FO}^{k=N}[\,]$ is closed under preimages under homomorphisms.*

Finally:

Proposition A.4. *For every natural number N, the class $\mathbf{FO}^{k=N}[\,]$ is a variety of languages.*

A.3 Classes $\mathbf{FO}^k[\mathsf{s}]$ are varieties of languages

In the last subsection of this appendix, we prove that for each natural number $k \geq 2$, the class $\mathbf{FO}^k[\mathsf{s}]$ of languages that are definable with the successor relation and with only k variables is a variety of languages. The reason why we do not consider the cases where $k < 1$ is that the successor relation is relevant when at least two variables are allowed in the formulae ($s(x, x)$ being never satisfied).

Once again, $\mathbf{FO}^k[\mathsf{s}]$ is naturally closed under Boolean combinations. Hence, it remains to prove that it is closed under quotients and preimages under homomorphisms. In contrast with the previous subsection, our proof of these closure properties does not rely on a semantic characterisation of $\mathbf{FO}^k[\mathsf{s}]$, but on a study of the equivalence relation $\equiv^k_d$. In this subsection, if v, w are two finite words over an alphabet $\mathbb{A}$, $v \equiv^k_d w$ denotes the fact that Duplicator has a winning strategy for the game $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$ defined on page 36: it is the Ehrenfeucht-Fraïssé game over v and w in d turns and with k tokens. In this game, v and w are seen as models of the signature $\mathbb{A} \sqcup \{\mathsf{s}\}$.

As we stated it in Subsection 1.2.3, the relation $\equiv^k_d$ is an equivalence relation over $\mathbb{A}^+$. In order to prove that $\mathbf{FO}^k[\mathsf{s}]$ satisfies the wanted closure properties, a key result is that the number of equivalence classes of $\equiv^k_d$ is finite and that each class is definable in $\mathrm{FO}^k_d[\mathbb{A}, \mathsf{s}]$, the set of $\mathrm{FO}^k[\mathbb{A}, \mathsf{s}]$ formulae of quantifier-depth at most d. To prove this, our first step is to show that the number of $\mathrm{FO}^k_d[\mathbb{A}, \mathsf{s}]$ formulae is finite, up to equivalence, as stated in the next lemma.

In this lemma, the k variables that are allowed in the formulae are, as usual, x_0, x_1,..., x_{k-1}. By $\bar{x}$ we denote any (possibly empty) subset of $\{x_0, \ldots, x_{k-1}\}$, and by $\varphi(\bar{x})$ we denote a formula whose set $FreeVar(\varphi)$ of free variables is the said subset $\bar{x}$. Hence, φ is a sentence if $\bar{x}$ is empty. Finally, we say that two such formulae $\varphi(\bar{x})$ and $\psi(\bar{x})$ are *equivalent* if for every word w in $\mathbb{A}^+$ and every valuation $\rho\colon \bar{x} \to \mathrm{Dom}(w)$, $w, \rho \vDash \varphi(\bar{x})$ if and only if $w, \rho \vDash \psi(\bar{x})$.

Lemma A.5. *Let $\mathbb{A}$ be an alphabet, let k and d be two natural numbers, with $k \geq 2$, and let $\bar{x}$ be a tuple of variables in $\{x_0, \ldots, x_{k-1}\}$. Then there exists a finite number of $\mathrm{FO}^k_d[\mathbb{A}, \mathsf{s}]$ formulae $\varphi^{k,d,\bar{x}}_0(\bar{x})$, $\varphi^{k,d,\bar{x}}_1(\bar{x})$,…, $\varphi^{k,d,\bar{x}}_{n-1}(\bar{x})$, all having $\bar{x}$ as free variables, such that every $\mathrm{FO}^k_d[\mathbb{A}, \mathsf{s}]$ formula $\varphi(\bar{x})$, also having $\bar{x}$ as free variables, is equivalent to one of the $\varphi^{k,d,\bar{x}}_\ell(\bar{x})$'s.*

This classic lemma can be generalised to model theory, when the involved signature is both finite and does not contain functional symbols of positive arity. Notice that the number n of such formulae depends on k, d, and $\bar{x}$, so we should technically write it $n_{k,d,\bar{x}}$. Yet the notation $\varphi^{k,d,\bar{x}}_{n_{k,d,\bar{x}}-1}(\bar{x})$ would become too heavy. In the end, denoting this natural number simply by n will not be problematic, since we are not interested in its actual value.

Proof. The proof goes by induction on the natural number d.

In the case when $d = 0$, the only atomic $\mathrm{FO}^k[\mathbb{A}, \mathsf{s}]$ formulae are the formulae of the shape $a(x_i)$, with $a \in \mathbb{A}$ and $i \in k$, of the shape $\mathsf{s}(x_i, x_j)$, or of the shape $x_i = x_j$, with $i, j \in k$. Since $\mathbb{A}$ is finite, there are only a finite number of them. A formula of depth 0 is a Boolean combination of these, and, since, up to equivalence, there is a finite number of Boolean combinations, there are, up to equivalence, a finite number of such formulae.

Now, let us suppose that the assertion is true for d: for every $\bar{x}' \subseteq \{x_0, \ldots, x_{k-1}\}$, there exists a finite number of $\mathrm{FO}^k_d[\mathbb{A}, \mathsf{s}]$ formulae $\varphi^{k,d,\bar{x}'}_0(\bar{x})$, $\varphi^{k,d,\bar{x}'}_1(\bar{x})$,…, $\varphi^{k,d,\bar{x}'}_{n-1}(\bar{x})$ such that every $\mathrm{FO}^k_d[\mathbb{A}, \mathsf{s}]$ formula $\varphi(\bar{x}')$ is equivalent to one of these.

We prove the assertion for $d+1$. Let $\bar{x} \subseteq \{x_0, \ldots, x_{k-1}\}$. Every $\mathrm{FO}^k_{d+1}[\mathbb{A}, \mathsf{s}]$ formula $\varphi(\bar{x})$ is a Boolean combination of formulae of the shape $Q_i x_i.\ \psi(\bar{x}')$, where: $i \in k$, Q_i is one of the two quantifiers $\forall$ and $\exists$, $\bar{x}' \subseteq \bar{x} \cup \{x_i\}$, and $\psi(\bar{x}') \in \mathrm{FO}^k_d[\mathbb{A}, \mathsf{s}]$. By assumption, $\psi(\bar{x}')$ is equivalent to one of the formulae $\varphi^{k,d,\bar{x}'}_\ell$'s, and, since there are only two distinct quantifiers, and, here again, since, up to equivalence, there is a finite number of Boolean combinations, there are, up to equivalence, a finite number of such formulae. $\square$

Notice that Lemma A.5 and our proof of it are not constructible: we do not give an actual value of n, and we do not provide a list of the formulae $\varphi^{k,d,\bar{x}}_0(\bar{x})$, $\varphi^{k,d,\bar{x}}_1(\bar{x})$,…, $\varphi^{k,d,\bar{x}}_{n-1}(\bar{x})$, nor we propose an algorithm that decides which of these a given $\mathrm{FO}^k_d[\mathbb{A}, \mathsf{s}]$ formula $\varphi(\bar{x})$ is equivalent to. We are only interested in the actual existence of these objects, which will be useful to prove the very next assertion.

We defined on page 22 the notion of equivalence classes. Here, in the context of $\equiv_d^k$, the equivalence class of a finite word $w \in \mathbb{A}^+$ is the set $[w]_d^k := \{v \in \mathbb{A}^+ \mid v \equiv_d^k w\}$. It is defined by a formula φ_{def}^w if $[w]_d^k = \mathcal{L}(\varphi_{\text{def}}^w)$.

Corollary A.6. *Let k and d two natural numbers, with $k \geq 2$. Then the equivalence relation $\equiv_d^k$ admits a finite number of equivalence classes, each of them being definable by an $\mathrm{FO}_d^k[\mathbb{A}, \mathsf{s}]$ formula.*

Proof. By Theorem 1.9 on page 37, each class $[w]_d^k$ can be characterised by the $\mathrm{FO}_d^k[\mathbb{A}, \mathsf{s}]$ sentences satisfied by w: let $\varphi_0^{k,d}$, $\varphi_1^{k,d}$,$\ldots$, $\varphi_{n-1}^{k,d}$ be the sentences obtained from Lemma A.5, meaning that every $\mathrm{FO}_d^k[\mathbb{A}, \mathsf{s}]$ sentence is equivalent to one of these. Then, for each word w in $\mathbb{A}^+$, we define I_w as the finite set $\left\{\ell \in n \mid w \vDash \varphi_\ell^{k,d}\right\}$.

By Theorem 1.9, $[w]_d^k$ is defined by the formula $\bigwedge_{\ell \in I_w} \varphi_\ell^{k,d} \wedge \bigwedge_{\ell \in n \setminus I_w} \neg\varphi_\ell^{k,d}$. Since there exists a finite number of subsets of $\{0, \ldots, n-1\}$, there exists a finite number of these formulae, and we have our result. $\square$

We can now deduce that $\mathbf{FO}_d^k[\mathbb{A}, \mathsf{s}]$ is exactly the class of languages that are equal to the union of the equivalence classes of their words:

Corollary A.7. *Let $L \subseteq \mathbb{A}^+$ be a language of finite words over an alphabet $\mathbb{A}$, and let k, d be two natural numbers, with $k \geq 2$. Then L is definable in $\mathrm{FO}_d^k[\mathbb{A}, \mathsf{s}]$ if and only if $L = \bigcup_{w \in L} [w]_d^k$.*

Proof. The first implication comes directly from Theorem 1.9 on page 37. Indeed, let us suppose that L is defined by some $\mathrm{FO}_d^k[\mathbb{A}, \mathsf{s}]$ formula φ, and let us show that the equality holds.

The inclusion $L \subseteq \bigcup_{w \in L} [w]_d^k$ is immediate: each word w in L belongs to its own class $[w]_d^k$. Now, let $w \in L$, and $v \in [w]_d^k$. Since $v \equiv_d^k w$, Theorem 1.9 tells us that v and w satisfy the same $\mathrm{FO}_d^k[\mathbb{A}, \mathsf{s}]$ sentences. In particular $v \vDash \varphi$, and therefore $v \in L$.

It is the other implication that actually makes use of Corollary A.6. Let us suppose that $L = \bigcup_{w \in L} [w]_d^k$. We know from Corollary A.6 that for each $w \in \mathbb{A}^+$, the particular class $[w]_d^k$ is definable by a formula φ_{def}^w. We define C_L as the set of these formulae φ_{def}^w, with $w \in L$. The same corollary tells us that C_L is finite, and therefore L is defined by the formula $\bigvee_{\varphi_{\text{def}}^w \in C_L} \varphi_{\text{def}}^w$: L is in $\mathbf{FO}_d^k[\mathbb{A}, \mathsf{s}]$, and we can conclude. $\square$

Now that we have proven this important proposition, we show that $\equiv_d^k$ admits convenient closure properties, which we will relate to the closure properties of $\mathbf{FO}^k[\mathbb{A}]$:

Proposition A.8. *Let $v, w \in \mathbb{A}^+$, and k, d be two natural numbers, with $k \geq 2$. Let $u \in \mathbb{A}^+$, and let h be a homomorphism from $\mathbb{A}^+$ to $\mathbb{B}^+$. If $v \equiv_d^k w$, then the three equivalences $u \cdot v \equiv_d^k u \cdot w$, $v \cdot u \equiv_d^k w \cdot u$, and $h(v) \equiv_d^k h(w)$ also hold.*

Proof. We shortly recall how the game $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$ is played. There are $2k$ tokens: two with the number 0 on them, two with the number 1 on them, and so on up to the number $k-1$. On each turn, Spoiler places one token on a position of one of the two words, and Duplicator answers by placing the corresponding token on a position of the other word. At the end of each of the d turns, the two players have defined a tuple $(x_i)_i$ of positions of v and another tuple $(y_i)_i$ of positions of w, both indexed by the numbers on the placed tokens. Duplicator wins the game if at each turn, the following holds for all i, j: $v(x_i) = w(y_i)$, $x_i = x_j$ if and only if $y_i = y_j$, and $x_i = x_j+1$ if and only if $y_i = y_j+1$.

So, let us suppose that $v \equiv_d^k w$, meaning that Duplicator has a strategy to win this game no matter the choices of Spoiler. We set $m = |v|$ and $n = |w|$.

We begin to show the second equivalence $v \cdot u \equiv_d^k w \cdot u$. By symmetry, we will be able to deduce the first equivalence $u \cdot v \equiv_d^k u \cdot w$.

To prove this, it is enough to show that $v \cdot a \equiv_d^k w \cdot a$ for any letter $a \in \mathbb{A}$: by repeating the procedure $|u|$ times, we obtain a proof of $v \cdot u \equiv_d^k w \cdot u$.

First, let us notice that, during any turn but the last one of the game $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$, if Spoiler places a token i on the last position of v (resp. the last position of w), then, in order to win, Duplicator has no choice but to place the other token i on the last position of w (resp. the last position of v). Indeed, if he does not, and places its token i on any other position $y_i < n-1$ of w, then on the next turn, Spoiler can place any other token j on the position $y_j = y_i+1$ of w, and Duplicator will have no place on v for his token j.

So, considering it, the winning strategy of Duplicator for the game $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v \cdot a, w \cdot a\rangle$ is rather simple to describe. Whenever Spoiler places a token i on a position $x_i < m$ of $v \cdot a$, meaning on a position of the original word v, Duplicator ignores the tokens placed on the last labelled-by-a positions, and follows its winning strategy in $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$. And symmetrically if Spoiler places a token i on a position $y_i < n$ of $w \cdot a$. Finally, Duplicator answers with position n to tokens on position m and reciprocally.

Let $\big(\langle x_i, y_i\rangle\big)_i$ be the pairs obtained at the end of one of the turns, meaning $(x_i)_i$ is a family of positions of $v \cdot a$, and $(y_i)_i$ is the corresponding family of positions of $w \cdot a$. Considering the winning strategy of $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$, the labels are preserved in the new pairs, and so is the successor relation for positions $x_i, x_j < m$, and $y_i, y_j < m$. It remains to make sure that if $x_i = m-1$, $x_j = m$, and $y_j = n$, then $y_i = n-1$, and, reciprocally, if

$y_i = n{-}1$, $y_j = n$, and $x_j = m$, then $x_i = m{-}1$. Notice that if any of the tokens during the game was placed on one of the additional labelled-by-a positions, then the respective token was placed on the other labelled-by-a, and this took one turn of the game. Therefore, the corresponding play in $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$ is at most $d{-}1$ turns long and therefore we can apply the remark we wrote in the previous paragraph, about the choices of the last positions in the strategy of $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$.

Figure 7 depicts this current situation over an example.

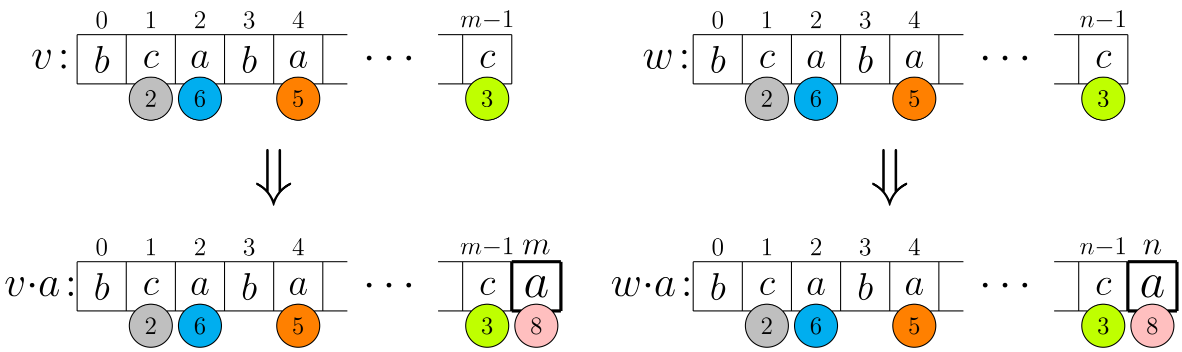


Figure 7: Duplicator's strategy on $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v \cdot a, w \cdot a\rangle$ is naturally induced from his strategy on $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$.

Now, it remains to prove that $h(v) \equiv_d^k h(w)$. Let us recall that, h being a homomorphism, we have the equality $h(v) = h\big(v(0)\big) \cdot h\big(v(1)\big) \cdots h\big(v(|v|{-}1)\big)$. This highlights a natural correspondence between positions of $h(v)$ and positions of v: for each position $x \in \mathrm{Dom}(h(v))$, we define $f_v(x)$ as the least position z of v such that $x \leq \Sigma_{z' \leq z} |h\big(v(z')\big)|$.

The function f is *increasing* (if $x \leq x'$ then $f_v(x) \leq f_v(x')$), but it is *a priori* not injective, since some letters of v might be mapped by h to words that are not single letters. In fact, for each position z of $\mathrm{Dom}(v)$, there are exactly $|h\big(v(z)\big)|$ positions x of $h(v)$ such that $f_v(x) = z$. We define g_v as the function that maps the first of them to 0, the second of them to 1, and so on up to $|h\big(v(z)\big)|{-}1$. The position mapped to 0 is called *minimal*, while the position mapped to $|h\big(v(z)\big)|{-}1$ is called *maximal*.

Figure 8 depicts an example and make these definitions more visual. Here, $\mathbb{A}$ is the alphabet $\{a_0, a_1, a_2\}$, $\mathbb{B}$ is the alphabet $\{b_0, b_1, b_2\}$, h is the homomorphism that maps a_0 to the word $b_0{\cdot}b_1{\cdot}b_1$, a_1 to the single-letter word b_2, and a_2 to the word $b_0{\cdot}b_0$, and, finally, v is the word $a_0{\cdot}a_1{\cdot}a_2$. With these settings, f_v maps the positions 0, 1, and 2 of $h(v)$ to the position 0 of v, the position 3 to the position 1, and the positions 4 and 5 to 2. The function g_v maps 0 to 1, 1 to 1, 2 to 2, 3 to 0, 4 to 0, and 5 to 1. The positions 0, 3, and 4 are minimal, while the positions 2, 3, and 5 are maximal.

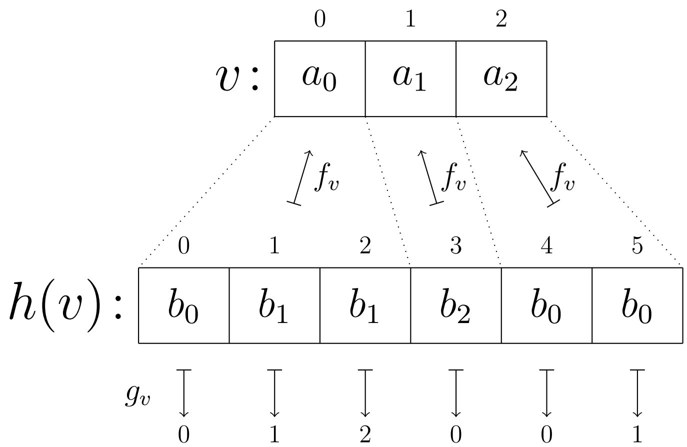


Figure 8: The definition of the functions f_v and g_v, over an example.

We define similar functions f_w and g_w for the word w.

Before finally describing Duplicator's strategy for $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{B},\mathsf{s}]}\langle h(v), h(w)\rangle$, we remark the following important equivalence: for all positions x_i, x_j of $h(v)$, we have $x_j = x_i+1$ if and only if:

– either $f_v(x_i) = f_v(x_j)$ and $g_v(x_j) = g_v(x_i) + 1$,

– or $f_v(x_j) = f_v(x_i) + 1$, x_i is maximal, and x_j is minimal.

We have naturally a similar equivalence for positions of $h(w)$.

Now, we explain how Duplicator wins $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{B},\mathsf{s}]}\langle h(v), h(w)\rangle$. During the game, he places in parallel tokens on words v and w, in order to make use of his winning strategy for the game $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$. When Spoiler places some token i on a position x_i of $h(v)$, Duplicator places a token i on the position $z_i = f_v(x_i)$ of v. From this, his strategy for $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$ tells him to place a token on some position z'_i of w. He places this token on z'_i, and comes back to $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{B},\mathsf{s}]}\langle h(v), h(w)\rangle$ by placing his second token i on the unique position y_i of $h(w)$ that satisfies $f_w(y_i) = z'_i$ and $g_w(y_i) = g_v(x_i)$. Duplicator plays symmetrically if Spoiler places a token on a position y_i of $h(w)$.

First, by the definitions of the functions f_v, f_w, g_v, and g_w, the strategy naturally preserves the labels and the equalities, since the strategy $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$ does. Now, considering the equivalence written above, the strategy also preserves the successor relation.

Indeed, let us suppose that, after some number of turns, two positions x_i and x_j of $h(v)$ have been defined with tokens, and that they satisfy $x_j = x_i+1$. This equality translates into one of the two conditions over v above. Let us suppose for instance that $f_v(x_j) = f_v(x_i)+1$, x_i is maximal, and x_j is minimal. Since Duplicator's strategy for $\mathcal{G}_d^{\mathrm{FO}^k[\mathbb{A},\mathsf{s}]}\langle v, w\rangle$ is winning, we also have $f_w(y_j) = f_w(y_i) + 1$. Also, by the definition of g_v and g_w, y_i is maximal, and y_j is minimal. Hence, $y_j = y_i + 1$. In the case when $f_w(y_i) = f_w(y_j)$ and $g_v(x_j) = g_v(x_i) + 1$, we would have come to the same conclusion.

Therefore, the strategy which we proposed for Duplicator is winning, and we obtain the relation $h(v) \equiv_d^k h(w)$. We can conclude the proof. $\square$

From this proposition, we deduce that $\mathbf{FO}^k[\mathsf{s}]$ is closed under quotients, as well as under preimages under homomorphisms, and we finally conclude this appendix:

Proposition A.9. *For every natural number $k \geq 2$, the class $\mathbf{FO}^k[\mathsf{s}]$ is a variety of languages.*

Proof. Let $k \geq 2$, let $L \in \mathbf{FO}^k[\mathbb{A}, \mathsf{s}]$, defined by some $\mathrm{FO}^k[\mathsf{s}]$ formula φ, and let d be its quantifier-depth. Let $u \in \mathbb{A}^+$. We show that $u^{-1} \cdot L = \bigcup_{w \in u^{-1} \cdot L} [w]_d^k$. Once again, the inclusion $u^{-1} \cdot L \subseteq \bigcup_{w \in u^{-1} \cdot L} [w]_d^k$ is immediate.

Let $v \in [w]_d^k$, where w is some word in $u^{-1} \cdot L$. We have $v \equiv_d^k w$ by definition and therefore $u \cdot v \equiv_d^k u \cdot w$ by Proposition A.8. Since $u \cdot w \vDash \varphi$, and φ is in $\mathrm{FO}_d^k[\mathbb{A}, \mathsf{s}]$, we know from Theorem 1.9 that $u \cdot v \vDash \varphi$. Thus, $v \in u^{-1} \cdot L$ and we can deduce the equality $u^{-1} \cdot L = \bigcup_{w \in u^{-1} \cdot L} [w]_d^k$.

Now, Corollary A.7 tells us that $u^{-1} \cdot L$ is definable in $\mathrm{FO}_d^k[\mathbb{A}, \mathsf{s}]$. Symmetrically, the second quotient language $L \cdot u^{-1}$ is also in $\mathbf{FO}_d^k[\mathbb{A}, \mathsf{s}]$, and we deduce that $\mathbf{FO}^k[\mathbb{A}, \mathsf{s}]$ is closed under quotients.

The same way, we use Proposition A.8 to prove that, if h is a homomorphism from $\mathbb{A}^+$ to $\mathbb{B}^+$, and if $L \in \mathbf{FO}_d^k[\mathbb{B}, \mathsf{s}]$, then $h^{-1}(L) = \bigcup_{w \in h^{-1}(L)} [w]_d^k$, and we deduce with Corollary A.7 that $h^{-1}(L) \in \mathbf{FO}^k[\mathbb{A}, \mathsf{s}]$. Thus, we conclude: $\mathbf{FO}^k[\mathsf{s}]$ is closed under preimages under homomorphisms, and it is a variety of languages. $\square$

As explained previously, our proof of Proposition A.9 is not constructive: we do not provide an algorithm that inputs a word $u \in \mathbb{A}^+$, a formula defining a language $L \subseteq \mathbb{A}^+$, and returns formulae defining the quotient languages $u^{-1} \cdot L$ and $L \cdot u^{-1}$. Nor do we provide an algorithm that, given a homomorphism $h \colon \mathbb{A}^+ \to \mathbb{B}^+$ and a formula defining a language $L \subseteq \mathbb{B}^+$, returns a formula defining its preimage $h^1(L)$.

B On projection of functional relations in $\mathbf{FO}^2[<]$

In this appendix, we provide a proof of Proposition 2.34, stating, in the conclusion of Chapter 2, that the projection of a functional relation in $\mathbf{FO}^2[<]$ is also in $\mathbf{FO}^2[<]$:

Proposition 2.34. *Let $\mathbb{A}$ and $\mathbb{B}$ be two alphabets, and let $F \in \mathbf{FO}^2[\binom{\mathbb{A}}{\mathbb{B}}, <]$. If F is functional, then its projection $\Pi_{\mathbb{A}^+}(F)$ is in $\mathbf{FO}^2[\mathbb{A}, <]$.*

To prove this proposition, we write down one among many semantic characterisations of $\mathbf{FO}^2[<]$.

An *unambiguous monomial* over an alphabet $\mathbb{A}$ is a language $L \subseteq \mathbb{A}^+$ that can be written as $\mathbb{A}_0^* \cdot a_0 \cdot \mathbb{A}_1^* \cdots \mathbb{A}_{n-1}^* \cdot a_{n-1} \cdot \mathbb{A}_n^* \setminus \{\epsilon\}$, where:

- each a_i is a letter of $\mathbb{A}$;
- each $\mathbb{A}_i$ is a subset of $\mathbb{A}$;
- any word w in L admits a unique decomposition $w = w_0 \cdot a_0 \cdot w_1 \cdots w_{n-1} \cdot a_{n-1} \cdot w_n$, with each w_i being in $\mathbb{A}_i^*$.

It is important to notice that this notion of unambiguous monomials does not exactly coincide with the notion of monomials which we defined in Section 3.4.

A language of finite words over $\mathbb{A}$ is called *unambiguous* if it is a finite union of disjoint unambiguous monomials over $\mathbb{A}$. The article [PW97] proves that the class of formulae $\mathrm{FO}^2[<]$ is capable of defining exactly these unambiguous languages:

Theorem B.1. *Let $L \subseteq \mathbb{A}^+$ be a language of finite words, over an alphabet $\mathbb{A}$. Then L is definable in $\mathrm{FO}^2[<]$ if and only if it is unambiguous.*

In order to prove Proposition 2.34, we prove the following lemma:

Lemma B.2. *Let $F \subseteq \binom{\mathbb{A}}{\mathbb{B}}^+$ be a binary relation between words over $\mathbb{A}$ and $\mathbb{B}$. We suppose moreover that it is functional and that it is an unambiguous monomial over $\binom{\mathbb{A}}{\mathbb{B}}$. Then the projection of F onto $\mathbb{A}^+$ is an unambiguous monomial over $\mathbb{A}$.*

Proof. We write F as $\mathbb{C}_0^* \cdot \binom{a_0}{b_0} \cdot \mathbb{C}_1^* \cdots \mathbb{C}_{n-1}^* \cdot \binom{a_{n-1}}{b_{n-1}} \cdot \mathbb{C}_n^* \setminus \{\epsilon\}$, where each a_i is a letter of $\mathbb{A}$, each b_i is a letter of $\mathbb{B}$, and each $\mathbb{C}_i$ is a subset of the product alphabet $\binom{\mathbb{A}}{\mathbb{B}}$.

First, it is clear that, if, for all $i \in n{+}1$, we write $\mathbb{A}_i$ for $\Pi_{\mathbb{A}}(\mathbb{C}_i)$, meaning the subset $\{a \in \mathbb{A} \mid \text{there exists some } b \in \mathbb{B} \text{ such that } \binom{a}{b} \in \mathbb{C}_i\} \subseteq \mathbb{A}$, then $\Pi_{\mathbb{A}^+}(F)$ is exactly the language $L = \mathbb{A}_0^* \cdot a_0 \cdot \mathbb{A}_1^* \cdots \mathbb{A}_{n-1}^* \cdot a_{n-1} \cdot \mathbb{A}_n^* \setminus \{\epsilon\}$. It remains to show that this monomial is unambiguous.

This is what we do now. Let $w \in L$, and let us decompose it in two different ways, in order to prove that these decompositions are in fact the same: $w = w_0 \cdot a_0 \cdot w_1 \cdots w_{n-1} \cdot a_{n-1} \cdot w_n$, and $w = w'_0 \cdot a_0 \cdot w'_1 \cdots w'_{n-1} \cdot a_{n-1} \cdot w'_n$, with w_i and w'_i being words in $\mathbb{A}_i^*$.

Let $i \in n{+}1$. For each position $x \in \mathrm{Dom}(w_i)$, there exists a letter $b \in \mathbb{B}$ such that $\binom{w_i(x)}{b} \in \mathbb{C}_i$. By naming $\sigma_i(x)$ this letter, we obtain a word $\sigma_i \in \mathbb{B}^{\mathrm{Dom}(w_i)}$ satisfying $\binom{w_i}{\sigma_i} \in \mathbb{C}_i^*$, and by defining σ as the word $\sigma_0{\cdot}b_0{\cdot}\sigma_1 \cdots \sigma_{n-1}{\cdot}b_{n-1}{\cdot}\sigma_n$, we get $\binom{w}{\sigma} \in F$.

The same way, we define words σ'_i such that for all $i \in n{+}1$, $\binom{w'_i}{\sigma'_i}$ is in $\mathbb{C}_i$, and we define σ' as the word $\sigma'_0{\cdot}b_0{\cdot}\sigma'_1 \cdots \sigma'_{n-1}{\cdot}b_{n-1}{\cdot}\sigma'_n$, which also satisfies $\binom{w}{\sigma'} \in F$.

Since F is functional, σ and σ' are necessarily the same words, and the same for the words $\binom{w}{\sigma}$ and $\binom{w}{\sigma'}$:

$$\begin{aligned} &\begin{pmatrix} w_0 \\ \sigma_0 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \sigma_1 \end{pmatrix} \cdots \begin{pmatrix} w_{n-1} \\ \sigma_{n-1} \end{pmatrix} \cdot \begin{pmatrix} a_{n-1} \\ b_{n-1} \end{pmatrix} \cdot \begin{pmatrix} w_n \\ \sigma_n \end{pmatrix} \\ = &\begin{pmatrix} w'_0 \\ \sigma'_0 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} \cdot \begin{pmatrix} w'_1 \\ \sigma'_1 \end{pmatrix} \cdots \begin{pmatrix} w'_{n-1} \\ \sigma'_{n-1} \end{pmatrix} \cdot \begin{pmatrix} a_{n-1} \\ b_{n-1} \end{pmatrix} \cdot \begin{pmatrix} w'_n \\ \sigma'_n \end{pmatrix}. \end{aligned}$$

By the unambiguity of F, this equality implies that for every $i \in n{+}1$, $\binom{w_i}{\sigma_i}$ and $\binom{w'_i}{\sigma'_i}$ are the same words, and, in particular, $w_i = w'_i$. Hence, we have proven the unicity of the decomposition of w: $\Pi_{\mathbb{A}^+}(F)$ is an unambiguous monomial. $\square$

Now, we can use Theorem B.1 and Lemma B.2 to finally prove Proposition 2.34:

Proof of Proposition 2.34. By Theorem B.1, we can write F as a disjoint union $\bigsqcup_{j \in p} F_j$, where each F_j is a functional unambiguous monomial over $\binom{\mathbb{A}}{\mathbb{B}}$.

By definition, we have the equality $\Pi_{\mathbb{A}^+}(F) = \bigcup_{j \in p} \Pi_{\mathbb{A}^+}(F_j)$. Hence, by Lemma B.2, F is a union of unambiguous monomials. It remains to prove that this union is disjoint.

We can conclude that via the remark which we wrote on page 79: the fact that F is functional and that the F_j's are pairwise disjoint ensures that their projections are pairwise disjoint too.

By Theorem B.1, we conclude that $\Pi_{\mathbb{A}^+}(F)$ is definable in $\mathrm{FO}^2[<]$. $\square$

C First-order choice functions for finitary linear orders

This appendix is devoted to the proof of the following proposition, given at the end of Section 4.6:

Proposition 4.34. *Let λ be a finitary linear order, whose construction does not involve the concatenation operation $+$, nor the η-operation. Then it admits a first-order choice function.*

We call *addition-free* such a finitary linear order, whose construction does not involve the operation $+$, nor the η-operation. It is isomorphic to a linear order of the form $\xi_0 \times \xi_1 \times \cdots \times \xi_{m-1}$, for some natural number m, where each ξ_i is either ω or $\omega^\star$. Also, we recall that a choice function over λ is called first order if it is defined by a formula $\varphi^\lambda_{\text{choice}}(X, x)$ that does not contain second-order quantifiers, see the definition on page 53.

Proposition 4.34 is unsatisfactory in itself, since very few finitary linear orders are addition-free. Yet, the tools used to construct the formula $\varphi^\lambda_{\text{choice}}(X, x)$ might be a good thing to start with, in order to prove Conjecture 4.33, stating on page 145, that any rigid finitary linear order admits a first-order choice function.

In this appendix, we prove in fact a stronger result, stating that all addition-free linear orders admit a first-order well order. Recall that a binary relation R over a linear order λ is defined by a formula $\varphi(x, y)$, with two first-order free variables, if R is the set of pairs $\langle x, y\rangle$ in λ^2 such that $\lambda \vDash \varphi(x, y)$. If said formula φ does not quantify over second-order variables, then R is called *first order*.

Proposition C.1. *Any addition-free linear order admits a first-order well order.*

Indeed, recall that if a first-order formula $\varphi^\lambda_{\text{wo}}(x, y)$ defines a well order over λ, then the first-order formula $\varphi^\lambda_{\text{choice}}(X, x) := x \in X \wedge \forall y{\in}X.\ \varphi^\lambda_{\text{wo}}(x, y)$ defines a choice function over λ.

The rest of this section is devoted to Proposition C.1. We begin by making sure that addition-free linear orders are necessarily rigid, meaning that they do not admit non-trivial automorphisms.

For convenience, we reason up to isomophism, meaning that we assume that each addition-free linear order λ *is* of the form $\xi_0 \times \xi_1 \times \cdots \times \xi_{m-1}$ above, the natural number m being called the *depth* of λ. Hence, each position of λ is an m-tuple of integers $\langle i_0, i_1, \ldots, i_{m-1}\rangle$, where, for each $k \in m$, the i_k's range over ξ_k.

For convenience, in this appendix, and only in this appendix, we choose to write the m-tuples in the reverse order, and to see $\omega^\star$ as the set $\mathbb{N} = \{0, 1, 2, \ldots\}$, but having its order reversed: $\ldots < 2 < 1 < 0$. This way, λ can be seen as $\mathbb{N}^m$, the set of words over the infinite alphabet $\mathbb{N}$ and of length m, whose lexicographic order corresponds to $\sqsubseteq^{\lambda}_{\text{lex}}$, the alternative lexicographic order of λ, defined on page 139. This alternative lexicographic order is a well order (see Proposition 4.29), and our point is to define it in FO[$<$].

Thus, in the order $\omega \times \omega$, the elements in the least copy of ω are $0{\cdot}0 < 0{\cdot}1 < 0{\cdot}2 < \ldots$; in the order $\omega^\star \times \omega^\star$, the elements in the greatest copy of $\omega^\star$ are $\ldots < 0{\cdot}2 < 0{\cdot}1 < 0{\cdot}0$.

As stated above, addition-free linear orders do not admit non-trivial automorphisms:

Fact C.1.1. *Any addition-free linear order is rigid.*

Proof. We only give a scheme of the proof, which can be seen as a generalisation of the proofs from Example 4.2 and Proposition 5.8, which state that the orders ω and $\omega^\star \times \omega$ are rigid. The proof has the following structure:

We consider any addition-free linear order $\lambda = \xi_0 \times \xi_1 \times \cdots \times \xi_{m-1}$, and any automorphism α of it.

First, we show by induction that for every $m \geq n \geq 0$, the elements of λ that are of the form $i_{m-1}{\cdot}\cdots{\cdot}i_n{\cdot}0{\cdot}\cdots{\cdot}0$ are necessarily mapped by α to elements of the same shape.

Then, in a second step, we prove, this time by a reverse induction, that those elements of the form $i_{m-1}{\cdot}\cdots{\cdot}i_n{\cdot}0{\cdot}\cdots{\cdot}0$ are necessarily mapped to themself. This assertion, in the case $n = 0$, concludes the proposition. $\square$

Now that we know that addition-free linear orders are rigid, our next step, before defining our alternative lexicographic order in FO[$<$], is to define every position. For instance, we saw in Example 4.3, on page 107, a way of defining in FO[$<$] each position of ω: the position n being "the position that has exactly n elements before itself". In a similar way, in the linear order $\omega \times \omega$, positions of the form $i{\cdot}0$, for some $i \in \omega$, can be defined as the elements that do not have *predecessors* (by defining here that x is a *predecessor* of y if y is a successor of x, in the sense that there is no third element between them).

To define these properties in the general case, we introduce a notion of *φ-predecessors* and *φ-successors.* Let $\varphi(x)$ be a first-order formula, that has exactly one free variable. We define $\mathsf{s}[\varphi](x, y)$ as the formula $x < y \wedge \varphi(y) \wedge \forall z{>}x.\ \varphi(z) \Rightarrow y \leq z$. If x and y are two elements of a linear order λ such that $\lambda \vDash \mathsf{s}[\varphi](x, y)$, then we say that y is the *φ-successor* of x. Notice that it does not involve x satisfying φ (in the sense that this does not imply $\lambda \vDash \varphi(x)$). Symmetrically, we define $\mathsf{p}[\varphi](x, y)$ as the formula $x < y \wedge \varphi(x) \wedge \forall z{<}y.\ \varphi(z) \Rightarrow z \leq x$. If $\lambda \vDash \mathsf{p}[\varphi](x, y)$, we say that x is the *φ-predecessor* of y. Once again, this does not say anything about the satisfaction of φ by y.

Now, let ℓ be a natural number, we define the first-order predicate $\mathsf{Min}^\ell[\varphi](x)$ inductively: $\mathsf{Min}^0[\varphi](x_0)$ is the formula $\varphi(x_0) \wedge \neg\exists x.\ \mathsf{p}[\varphi](x, x_0)$, and, if $\ell \geq 1$, then $\mathsf{Min}^\ell[\varphi](x_\ell)$ is the formula $\exists x_{\ell-1}.\mathsf{Min}^{\ell-1}[\varphi](x_{\ell-1}) \wedge \mathsf{s}[\varphi](x_{\ell-1}, x_\ell)$. We define the first-order formula $\mathsf{Max}^\ell[\varphi](x)$ symmetrically.

Finally, we inductively define a family of first-order formulae, which we call the *min-max-predicates*: the formula $x = x$ (that we write $\top(x)$) is a min-max-predicate of depth 0, and, if $\varphi(x)$ is a min-max-predicate of depth n, and ℓ is any natural number, then the formulae $\mathsf{Min}^\ell[\varphi](x)$ and $\mathsf{Max}^\ell[\varphi](x)$ are min-max-predicates of depth $n{+}1$.

As the following lemma states it, this notion of min-max-predicates is convenient to define the positions of addition-free linear orders.

Lemma C.2. *Let λ be an addition-free linear order of depth $m \in \mathbb{N}$, and let $n \in \mathbb{N}$ such that $m \geq n$. Then, for all $i_{n-1}, \ldots, i_0 \in \mathbb{N}$, there exists a formula $\varphi^{\lambda,n}_{\uparrow i_{n-1},\ldots,i_0}$ which is a min-max-predicate of depth n and that is such that for all $j_{m-1}, \ldots, j_0 \in \mathbb{N}$, we have the equivalence $\lambda \vDash \varphi^{\lambda,n}_{\uparrow i_{n-1},\ldots,i_0}(j_{m-1}\cdots j_n \cdot j_{n-1}\cdots j_0)$ if and only if $j_k = i_k$ for every k such that $n > k$.*

Proof. We prove this lemma by induction on n.

If $n = 0$, then the wanted condition becomes trivial, and therefore $\top(x)$, the min-max-predicate of depth 0, naturally fulfils it.

Let us suppose now that $m \geq n \geq 1$ and that the claim is true for $n{-}1$, meaning that, for all $i_{n-2}, \ldots, i_0 \in \mathbb{N}$, we have constructed the min-max-predicate $\varphi^{\lambda,n-1}_{\uparrow i_{n-2},\ldots,i_0}$ of depth $n{-}1$ that satisfies our wanted property. Our present goal is to construct the formula $\varphi^{\lambda,n}_{\uparrow i_{n-1},\ldots,i_0}$, for some $i_{n-1}, \ldots, i_0 \in \mathbb{N}$.

For the rest of the proof, we simply write ψ for $\varphi^{\lambda,n-1}_{\uparrow i_{n-2},\ldots,i_0}$, and $\xi_0 \times \xi_1 \times \cdots \times \xi_{m-1}$ for λ. Let us suppose for a moment that ξ_{n-1} is ω: the case when it is $\omega^\star$ will be considered at the end of the proof.

By the assumption about ψ, and the definition of the min-max-predicates, we notice that, for all natural numbers $j_{m-1}, \ldots, j_0$, we have $i_k = j_k$ for all $n{-}1 \geq k$ if and only if the element $x = j_{m-1}\cdots j_0$ satisfies $\mathsf{Min}^{i_{n-1}}[\psi]$. The argument goes by induction on i_{n-1}. For the sake of readability, we denote this inductive parameter by ℓ. Yet, the argument for the heredity is very similar to the one for the initiation, therefore we only fully redact the case when $\ell = 0$.

First, let us suppose that the equality $i_k = j_k$ indeed holds for every $n{-}1 \geq k$. Then, in particular, $i_k = j_k$ for every $n{-}2 \geq k$, which means that x satisfies ψ, by the assumption about this formula. In order to prove that x also satisfies $\mathsf{Min}^0[\psi]$, let $y < x$ also satisfying ψ, meaning of the form $k_{m-1}\cdots k_n{\cdot}k_{n-1}{\cdot}i_{n-2}\cdots i_0$, with $k_{m-1},\ldots,k_{n-1} \in \mathbb{N}$.

Because $y < x = j_{m-1}\cdots j_n{\cdot}0{\cdot}i_{n-2}\cdots i_0$, we necessarily have $k_{m-1}\cdots k_n < j_{m-1}\cdots j_n$, in the linear order $\xi_n \times \cdots \times \xi_{m-1}$. Therefore, the element $z = k_{m-1}\cdots k_n{\cdot}(k_{n-1}{+}1){\cdot}i_{n-2}\cdots i_0$ is such that $y < z < x$, and moreover it also satisfies ψ. We have proven that x cannot have any ψ-predecessor, and therefore it satisfies $\mathsf{Min}^0[\psi]$.

Reciprocally, let us suppose that x satisfies $\mathsf{Min}^0[\psi]$. In particular it satisfies ψ, and therefore we have $i_k = j_k$ for all $n{-}1 > k$. It remains to prove that $j_k = 0$, which is almost immediate. Indeed, if $0 < j_{n-1}$, then we can verity that $j_{m-1}\cdots j_n{\cdot}(j_{n-1}{-}1){\cdot}i_{n-2}\cdots i_0$ is the ψ-predecessor of x, which no longer can satisfy $\mathsf{Min}^0[\psi]$.

As stated above, we prove by induction on ℓ the equivalence in the general case. This justifies the definition of $\varphi^{\lambda,n}_{\uparrow i_{n-1},\ldots,i_0}(x)$ as the min-max-predicate $\mathsf{Min}^{i_{n-1}}[\varphi^{\lambda,n-1}_{\uparrow i_{n-2},\ldots,i_0}](x)$, of depth n. Remember that, for a sake of simplification, we supposed that ξ_{n-1} is ω. If ξ_{n-1} is in fact $\omega^\star$, then a similar reasoning would conclude that the formula $\mathsf{Max}^{i_{n-1}}[\varphi^{\lambda,n-1}_{\uparrow i_{n-2},\ldots,i_0}](x)$ fulfils our need. We conclude the proof. $\square$

The formulae in Lemma C.2 talk about the equality of the last coordinates of the elements of addition-free linear orders. Using them, we can define a formula that talks about the equality of the first coordinates:

Lemma C.3. *Let λ be an addition-free linear order of depth $m \in \mathbb{N}$, and let $m \geq n$. Then there exists a first-order formula $\varphi^{\lambda,n}_{\downarrow}(x,y)$, with two free variables, such that, for all natural numbers $i_{m-1},\ldots,i_0,j_{m-1},\ldots,j_0$, we have $\lambda \vDash \varphi^{\lambda,n}_{\downarrow}\big(i_{m-1}\cdots i_0, j_{m-1}\cdots j_0\big)$ if and only if $i_k = j_k$ for every $m > k \geq n$.*

Proof. To prove this lemma, we first construct, for each $m \geq n \geq 0$, an intermediate first-order formula $\varphi^{\lambda,n}_{\text{zeros}}(x_0,x)$, that has the property that for all elements x_0 and x of λ, we have $\lambda \vDash \varphi^{\lambda,n}_{\text{zeros}}(x_0,x)$ if and only if we have the following implication: if x_0 is of the shape $i_{m-1}\cdots i_n{\cdot}0\cdots 0$, with $i_{m-1},\ldots,i_n$ being natural numbers, then x is necessarily of the shape $i_{m-1}\cdots i_n{\cdot}j_{n-1}\cdots j_0$, with $j_{n-1},\ldots,j_0$ being any natural numbers.

Indeed, if such a formula exists, then $\varphi^{\lambda,n}_{\downarrow}(x,y)$ defined as

$$\exists x_0.\ \varphi^{\lambda,n}_{\uparrow 0,\ldots,0}(x_0) \wedge \varphi^{\lambda,n}_{\text{zeros}}(x_0,x) \wedge \varphi^{\lambda,n}_{\text{zeros}}(x_0,y)$$

fulfils our need. We construct $\varphi^{\lambda,n}_{\text{zeros}}(x_0,x)$ by induction on n.

For $n = 0$, the formula $\varphi^{\lambda,n}_{\text{zeros}}$ simply becomes $x_0 = x$, which is itself a first-order formula.

Suppose now that $n \geq 1$, and that we have constructed $\varphi_{\text{zeros}}^{\lambda,n-1}$ fulfilling our need. We write λ as $\xi_0 \times \xi_1 \times \cdots \times \xi_{m-1}$, and we suppose for a moment that ξ_{n-1} is ω. The case when it is $\omega^\star$ will be symmetrically covered later. Then, we define $\varphi_{\text{zeros}}^{\lambda,n}(x_0, x)$ as the formula

$$\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}(x_0) \implies \left[\varphi_{\text{zeros}}^{\lambda,n-1}(x_0, x) \vee \left(\mathsf{p}[\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}](x_0, x) \wedge \forall y_0.\ \neg\big(\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}(y_0) \wedge \varphi_{\text{zeros}}^{\lambda,n-1}(y_0, x)\big)\right)\right].$$

Let us show that if fulfils our need.

In a first step, let x_0 and x be some elements of λ, such that $\lambda \vDash \varphi_{\text{zeros}}^{\lambda,n}(x_0, x)$.

Let us suppose that x_0 is of the shape $i_{m-1}\cdot\cdots\cdot i_n\cdot 0\cdot\cdots\cdot 0$, with $i_{m-1}, \ldots, i_n$ being natural numbers. This means that it satisfies $\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}$. Hence, we must have $\lambda \vDash \varphi_{\text{zeros}}^{\lambda,n-1}(x_0, x)$, or $\lambda \vDash \mathsf{p}[\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}](x_0, x) \wedge \forall y_0.\ \neg\big(\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}(y_0) \wedge \varphi_{\text{zeros}}^{\lambda,n-1}(y_0, x)\big)$.

If $\lambda \vDash \varphi_{\text{zeros}}^{\lambda,n-1}(x_0, x)$, then it means that x is of the shape $i_{m-1}\cdot\cdots\cdot i_n\cdot 0\cdot j_{n-2}\cdot\cdots\cdot j_0$, so, in particular, it is of the shape $i_{m-1}\cdot\cdots\cdot i_n\cdot j_{n-1}\cdot\cdots\cdot j_0$, with $j_{n-1}, \ldots, j_0$ being natural numbers.

Let us assume now that we are in the second case: we have both $\lambda \vDash \mathsf{p}[\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}](x_0, x)$ and $\lambda \vDash \forall y_0.\ \neg\big(\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}(y_0) \wedge \varphi_{\text{zeros}}^{\lambda,n-1}(y_0, x)\big)$. We write x as $j_{m-1}\cdot\cdots\cdot j_0$, and we want to show that we have the equality $i_k = j_k$ for every $m > k \geq n$. A first thing to notice is that the second condition implies that j_{n-1} cannot be equal to 0. Indeed, if it was, then $j_{m-1}\cdot\cdots\cdot j_n\cdot 0\cdot 0\cdot\cdots\cdot 0$ would be an element y_0 such that $\lambda \vDash \varphi_{\uparrow 0,\ldots,0}^{\lambda,n}(y_0) \wedge \varphi_{\text{zeros}}^{\lambda,n-1}(y_0, x)$. Now, suppose, in order to get to a contradiction, that there exists some $m > k \geq n$ such that $i_k \neq j_k$, and let us define y_0 as the element $j_{m-1}\cdot\cdots\cdot j_n\cdot 0\cdot\cdots\cdot 0$. Since there exists some $k \geq n$ such that $i_k \neq j_k$, and since $x_0 < x = j_{m-1}\cdot\cdots\cdot j_0$ (because $\lambda \vDash \mathsf{p}[\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}](x_0, x)$), we have necessarily $x_0 < y_0$. Also, because $0 < j_{n-1}$, by the remark above, we have $y_0 < x$. Hence, y_0 is an element that satisfies both $x_0 < y_0 < x$ and $\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}$, which contradicts $\lambda \vDash \mathsf{p}[\varphi_{\uparrow 0,\ldots,0}^{\lambda,n}](x_0, x)$. Therefore, we necessarily have $i_k = j_k$ for every $m > k \geq n$, and we have proven our implication.

Now, we prove the other implication of the equivalence. Let x_0 and x be two elements of λ, such if x_0 is of form $i_{m-1}\cdot\cdots\cdot i_n\cdot 0\cdot\cdots\cdot 0$, then x is necessarily of the form $i_{m-1}\cdot\cdots\cdot i_n\cdot j_{n-1}\cdot\cdots\cdot j_0$, and let us show that we have $\lambda \vDash \varphi_{\text{zeros}}^{\lambda,n}(x_0, x)$.

Let us suppose that $\lambda \vDash \varphi_{\uparrow 0,\ldots,0}^{\lambda,n}(x_0)$. Then it means that x_0 is indeed of the form $i_{m-1}\cdot\cdots\cdot i_n\cdot 0\cdot\cdots\cdot 0$, and, therefore, by our assumption, $x = i_{m-1}\cdot\cdots\cdot i_n\cdot j_{n-1}\cdot\cdots\cdot j_0$. Then, there are two possible cases. Either $j_{n-1} = 0$, and in that case we have $\lambda \vDash \varphi_{\text{zeros}}^{\lambda,n-1}(x_0, x)$, or $j_{n-1} > 0$. In the latter case, there cannot be any y_0 such that $\lambda \vDash \varphi_{\uparrow 0,\ldots,0}^{\lambda,n}(y_0) \wedge \varphi_{\text{zeros}}^{\lambda,n-1}(y_0, x)$. Also, $x_0 < x$ and there cannot either be any element y_0 that satisfies both $x_0 < y_0 < x$ and $\lambda \vDash \varphi_{\uparrow 0,\ldots,0}^{\lambda,n}(y_0, x)$, which concludes the implication.

Hence, we have proven the equivalence: our formula $\varphi_{\text{zeros}}^{\lambda,n}(x_0, x)$ defines what we need.

Remember that we assumed that ξ_{n-1} is the linear order ω. If ξ_{n-1} is the linear order $\omega^\star$, then $\varphi^{\lambda,n}_{\text{zeros}}(x_0, x)$ is defined as the formula

$$\varphi^{\lambda,n}_{\uparrow 0,\dots,0}(x_0) \implies \Big[\varphi^{\lambda,n-1}_{\text{zeros}}(x_0, x) \vee \Big(\mathsf{s}[\varphi^{\lambda,n}_{\uparrow 0,\dots,0}](x, x_0) \wedge \forall y_0.\ \neg\big(\varphi^{\lambda,n}_{\uparrow 0,\dots,0}(y_0) \wedge \varphi^{\lambda,n-1}_{\text{zeros}}(y_0, x)\big)\Big)\Big].$$

We conclude the proof. $\square$

Now that we have constructed our formulae $\varphi^{\lambda,n}_{\downarrow}$, we can define the alternative lexicographic order of λ:

Proposition C.4. *Let λ be an addition-free linear order, then there exists an* FO[<] *formula $\varphi^{\lambda}_{\text{lex}}(x, y)$ that defines the alternative lexicographic order of λ.*

Proof. This proof is also produced by induction. More precisely, if m is the depth of λ, we prove that for each natural number n such that $m \geq n$, there exists a formula $\varphi^{\lambda,n}_{\text{lex}}(x, y)$ such that for all elements $x = i_{m-1} \cdots i_0, y = j_{m-1} \cdots j_0$ in λ, we have $\lambda \vDash \varphi^{\lambda,n}_{\text{lex}}(x, y)$ if and only if for all $m > k \geq n$, $i_k = j_k$, and moreover $x \sqsubseteq^{\lambda}_{\text{lex}} y$.

In the case $n = 0$, the condition becomes $x = y$, which is itself a first-order formula.

Now, let us suppose that the result is true for $n-1$, meaning that we have constructed a satisfactory first-order formula $\varphi^{\lambda,n-1}_{\text{lex}}(x, y)$. We write $\xi_0 \times \cdots \times \xi_{m-1}$ for λ. We suppose ξ_{n-1} to be ω, the case when it is $\omega^\star$ being symmetrically covered later. Then, we define $\varphi^{\lambda,n}_{\text{lex}}(x, y)$ as the formula

$$\begin{aligned}\varphi^{\lambda,n}_{\downarrow}(x, y) \wedge\ & \big(\varphi^{\lambda,n-1}_{\downarrow}(x, y) \implies \varphi^{\lambda,n-1}_{\text{lex}}(x, y)\big) \\ \wedge\ & \big(\neg\varphi^{\lambda,n-1}_{\downarrow}(x, y) \implies x < y\big).\end{aligned}$$

It can verified easily that this formula defines the right condition.

Indeed, let $x = i_{m-1} \cdots i_0$ and $y = j_{m-1} \cdots j_0$ be two elements of λ. By the definition of $\varphi^{\lambda,n}_{\downarrow}(x, y)$ in Lemma C.3, $\lambda \vDash \varphi^{\lambda,n}_{\downarrow}(x, y)$ is equivalent to $i_k = j_k$ for every $m > k \geq n$. Under that condition, $\lambda \vDash \varphi^{\lambda,n-1}_{\downarrow}(x, y)$ is equivalent to $i_{n-1} = j_{n-1}$. When it is not fulfilled, meaning if $i_{n-1} \neq j_{n-1}$, then $x \sqsubseteq^{\lambda}_{\text{lex}} y$ is equivalent to $x < y$.

In the case when ξ_{n-1} is $\omega^\star$, then $\varphi^{\lambda,n}_{\text{lex}}(x, y)$ is defined similarly, except $x < y$ in the last implication is substituted by $y < x$.

Now that we constructed $\varphi^{\lambda,n}_{\text{lex}}$ for all $n \leq m$, we can simply define $\varphi^{\lambda}_{\text{lex}}(x, y)$ as $\varphi^{\lambda,m}_{\text{lex}}(x, y)$, and the property is proven. $\square$

From Proposition C.4 we deduce Propositions C.1 and 4.34 and we can conclude this appendix.