



# Duality for the existential fragment of first-order logic on words

Mehdi Zaïdi

## ► To cite this version:

Mehdi Zaïdi. Duality for the existential fragment of first-order logic on words. General Topology [math.GN]. Université Côte d'Azur, 2022. English. NNT : 2022COAZ4064 . tel-03882130v2

**HAL Id: tel-03882130**

**<https://hal.science/tel-03882130v2>**

Submitted on 19 Dec 2022

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

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



$$\rho \left( \frac{\partial v}{\partial t} + v \cdot \nabla v \right) = -\nabla p + \nabla \cdot T + f$$

$$e^{i\pi} + 1 = 0$$

# THÈSE DE DOCTORAT

Dualité pour le fragment existentiel de  
la logique du premier ordre sur les  
mots

**Mehdi ZAÏDI**

Laboratoire de Mathématiques J.A Dieudonné (LJAD)

**Présentée en vue de l'obtention  
du grade de docteur en Mathématiques  
d'Université Côte d'Azur**

**Dirigée par :** Mai Gehrke

**Co-encadrée par :** Sam van Gool

**Soutenue le :** 7 Novembre 2022

**Devant le jury, composé de :**

Clemens BERGER, MCF, LJAD

Mai GEHRKE, PR, LJAD

Vincenzo MARRA, PR, Università di Milano

Paul-André MELLIÈS, PR, IRIF

Sam VAN GOOL, MCF, IRIF



---

# Thèse de Doctorat

---

## DUALITY FOR THE EXISTENTIAL FRAGMENT OF FIRST-ORDER LOGIC ON WORDS

---

Présentée et soutenue par :

Mehdi ZAIDI

En vue de l'obtention du grade de

Docteur en Sciences  
de l'Université Côte d'Azur

Discipline : Mathématiques  
Laboratoire de Mathématiques J.A. Dieudonné (LJAD)

---

Thèse dirigée par Mai Gehrke et co-encadrée par Sam van Gool

Date de soutenance : Lundi 7 Novembre 2022 - 14 h

Devant le jury, composé de :

Clemens BERGER	MCF	Laboratoire J.A. Dieudonné	Examineur
Mai GEHRKE	PR	Laboratoire J.A. Dieudonné	Directrice de thèse - Examinatrice
Vincenzo MARRA	PR	Università di Milano	Rapporteur - Examineur
Paul-André MELLIÈS	PR	IRIF	Rapporteur - Examineur
Sam VAN GOOL	MCF	IRIF	Co-encadrant - Examineur

---

**Résumé:** Cette thèse s'intéresse à l'application de méthodes de dualité topologique à des problèmes de l'informatique théorique. Un des objectifs finaux de cette démarche est l'obtention de résultats en théorie de la complexité, via l'étude d'objets topologiques caractérisant les différentes classes de complexité. La logique est ce qui est à l'interface entre ces deux domaines en apparence très éloignés, plus particulièrement un sous-domaine de la théorie des modèles finis : la logique sur les mots. Il est possible de donner une description de certaines classes de complexité comme des familles de langages, potentiellement non réguliers, sur un alphabet fini.

Très peu de résultats de dualité sont connus pour les fragments de la logique du premier ordre sur les mots décrits par des langages qui sortent du cadre régulier. Notre contribution est l'étude détaillée d'un tel fragment. Pour un entier  $k \geq 1$  fixé, nous considérons l'algèbre de Boole  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ . Celle-ci correspond au fragment de logique sur les mots consistant en les combinaisons Booléennes de propositions définies en utilisant un bloc d'au plus  $k$  quantificateurs existentiels, les prédicats sur les lettres et les prédicats numériques uniformes d'arité  $l \in \{1, \dots, k\}$ . Nous fournissons une étude détaillée de l'espace dual à cette algèbre de Boole, pour tout  $k \geq 1$ , et nous donnons plusieurs caractérisations de ses points. Dans le cas particulier où  $k = 1$ , nous sommes capables de construire une famille d'équations ultrafiltre qui caractérise l'algèbre de Boole  $\mathcal{B}\Sigma_1[\mathcal{N}_1^u]$ .

**Mots clefs :** Topologie générale, Dualité de Stone, Logique sur les mots, Théorie de la complexité descriptive, Treillis distributifs, Algèbre de Boole, Espace de Vietoris, Coloriages fini, Équations ultrafiltre.

**Abstract:** This thesis fits in the area of research that investigates the application of topological duality methods to problems that appear in theoretical computer science. One of the eventual goals of this approach is to derive results in computational complexity theory by studying appropriate topological objects which characterise them. The link which relates these two seemingly separated fields is logic, more precisely a subdomain of finite model theory known as logic on words. It allows for a description of complexity classes as certain families of languages, possibly non-regular, on a finite alphabet.

Very few is known about the duality theory relative to fragments of first-order logic on words which lie outside of the scope of regular languages. The contribution of our work is a detailed study of such a fragment. Fixing an integer  $k \geq 1$ , we consider the Boolean algebra  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ . It corresponds to the fragment of logic on words consisting in Boolean combinations of sentences defined by using a block of at most  $k$  existential quantifiers, letter predicates and uniform numerical predicates of arity  $l \in \{1, \dots, k\}$ . We give a detailed study of the dual space of this Boolean algebra, for any  $k \geq 1$ , and provide several characterisations of its points. In the particular case where  $k = 1$ , we are able to construct a family of ultrafilter equations which characterise the Boolean algebra  $\mathcal{B}\Sigma_1[\mathcal{N}_1^u]$ .

**Key words:** General topology, Stone duality, Logic on words, Descriptive complexity theory, Distributive lattices, Boolean algebra, Vietoris hyperspaces, Finite colourings, Ultrafilter equations



---

## Remerciements

---

Je m'excuse d'avance pour les individus que je n'ai pas volontairement oublié. Suivant la coutûme tacite, j'ordonne ces remerciements par degré d'implication dans cette thèse.

- Je remercie ma directrice de thèse Mai Gehrke, ainsi que mon co-encadrant Sam van Gool.
- Je remercie le projet ERC-DuaLL, le CNRS et le laboratoire Jean Dieudonné.
- Je remercie les relecteurs, Vincenzo Marra et Paul-André Melliès.
- Je remercie Luca Reggio, Tomas Jakl, Brett Mclean et Wesley Fussner.
- Je remercie le personnel de l'EDSFA, en particulier Catherine Briet et Elisabeth Taffin de Givenchy.
- Je remercie le personnel administratif du LJAD, en particulier Clara Salaun.
- Je remercie mon comité de suivi de thèse: Clemens Berger, Bruno Vallette et Emmanuel Militon
- Je remercie la société Arkeup, en particulier leurs tickets restaurant, ainsi que le Pôle Emploi.
- Pas vraiment un remerciement, parce que ce serait bizarre, mais S/O l'ensemble des entités que j'ai pu rencontrer par le biais de la thèse et du LJAD durant ces quelques années. Par ordre de rencontre: Thibault Lemoine, Frédéric Crozatier, Alexis Roquefeuil, Leonard Guetta, Axel Osmond, Benjamin Zarka, Léo Vivion, Yen (Tran Thi), Kévin Atsou, Julie Llobell, Reine Fares, Alexis Loyer, Jonathan Stéphano, Giulia Lissoni - Victor Marx, Yash Chopra, Eliot Pacherie - Alexis Gilles, Alexis Garcia, Yanis Mabed, Biao Ma, Hadrien Gascuel, Melodie Andrieu, Giulia Boulant, Armand Koenig, Marco Corneli, Luis Alberto Gomez Nava, Laurence Beaude, Billel Guelmame, Haroune Houamed, Marcella Bonazzoli, Tommy Wiseau, Frédérique Noel, Michaël Brunengo, Ali Aboudou Elarif, David Lafontaine, Samira Amraoui, Ludovick Gagnon, Thibaut Dumont, Zhixin Xie(Jay-Z), Nadine Dirani, Najwa Ghannoum, Angel Toledo, Leuss Porze, Cyril Karamaoun, Lali Hen, Alekos Cecchin, Riccardo Di Dio, Giulia Marchello, Cécile Gachet, Bochra Mejri, Thanos Vasileiadis, Dahmane Dechicha, Zakaria Ouaras, M'hammed Oudrane, Edouard Balzin, Gustave Billon.





<b>1</b>	<b>Duality and formal language theory</b>	<b>21</b>
1.1	Boolean algebras and Stone duality . . . . .	21
1.1.1	General duality theory . . . . .	22
1.1.2	The dual space of the powerset algebra . . . . .	27
1.2	Modal algebra and the Vietoris functor . . . . .	29
1.3	Formal languages and recognition . . . . .	36
1.4	Logic on words . . . . .	41
1.4.1	Syntax . . . . .	42
1.4.2	Semantics . . . . .	43
1.5	Ultrafilter equations . . . . .	45
<b>2</b>	<b>Duality for the existential fragment of first-order logic on words</b>	<b>49</b>
2.1	General setting . . . . .	51
2.2	Logical description of $\mathcal{B}_k$ . . . . .	54
2.3	The dual space via finite colourings . . . . .	57
2.3.1	Colourings approach . . . . .	57
2.3.2	Explicit characterisation of $X_1$ . . . . .	67
2.4	Projection of $X_k$ on $\mathcal{V}(\beta(\mathbb{N}^k))$ . . . . .	70
2.4.1	The Boolean algebra $\mathcal{B}_2^{a,a}$ and the interior operator $R$ . . . . .	70
2.4.2	Thick closed subsets . . . . .	76
<b>3</b>	<b>Ultrafilter equations for the existential fragment of logic on words</b>	<b>87</b>
3.1	An alternative presentation for $\mathcal{B}_k$ and $\mathcal{B}_{0,k}$ . . . . .	88
3.2	A certain family of ultrafilter equations . . . . .	90
3.2.1	General setting . . . . .	90
3.2.2	Ultrafilter equations in terms of finite colourings . . . . .	93
3.3	A topological proof for ultrafilter equations on $\mathcal{B}_{0,1}$ and $\mathcal{B}_1$ . . . . .	98
3.3.1	Soundness . . . . .	99
3.3.2	Completeness for $\mathcal{B}_1$ . . . . .	101
3.3.3	Completeness for $\mathcal{B}_{0,1}$ . . . . .	106



---

## List of Figures

---

1	Différentes logiques et leurs équivalents algébriques et topologiques . . . .	3
2	Some Stone type dualities . . . . .	12
3	Different logics and their algebraic and topological counterparts . . . . .	13
4	A few computational complexity classes . . . . .	13
5	Logic fragments associated to certain computational complexity classes . .	14
6	Eilenberg-Reiterman correspondence in the regular case . . . . .	16
7	The powerset algebra on $\{x, y, z\}$ , where $x, y$ and $z$ are pairwise disjoint elements. . . . .	23
8	Boolean Stone duality . . . . .	26
9	The space of ultrafilters $\beta(S)$ . . . . .	35
10	An automaton that recognises the language $A^*aA^*$ , where $A = \{a, b\}$ . . . .	37
11	A closed subset which is not in $X_2^{a,a}$ . . . . .	78
12	$\Delta^3$ . . . . .	79
13	A thick closed subset which is not in $X_2^{a,a}$ . . . . .	81

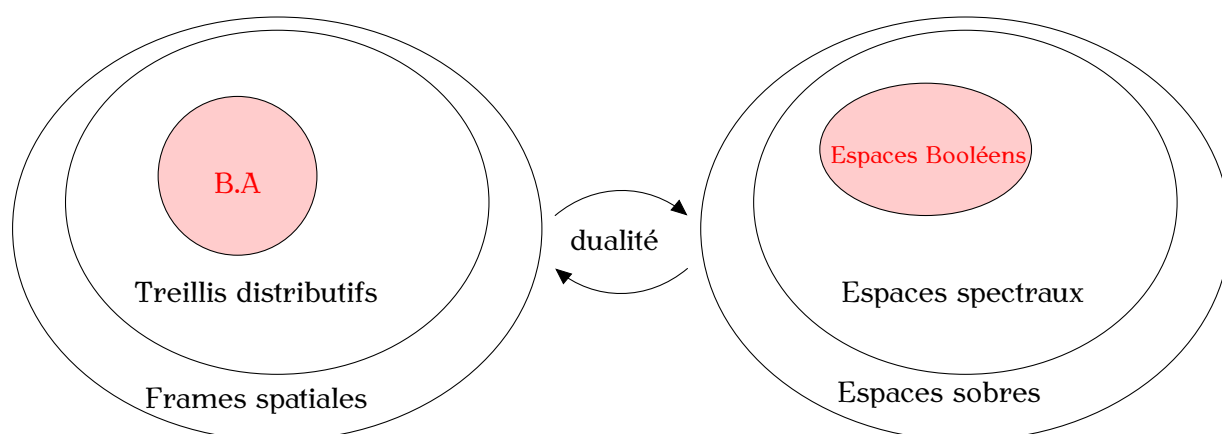


Cette thèse s'intéresse à l'application de méthodes de dualité topologique à des problèmes de l'informatique théorique. Un des objectifs finaux de cette démarche est l'obtention de résultats en théorie de la complexité, via l'étude d'objets topologiques caractérisant les différentes classes de complexité. La logique est ce qui est à l'interface entre ces deux domaines en apparence très éloignés, plus particulièrement un sous-domaine de la théorie des modèles finis : la logique sur les mots. Il est possible de donner une description de certaines classes de complexité comme des familles de langages, potentiellement non réguliers, sur un alphabet fini. Cette introduction donne un compte-rendu historique des travaux effectués dans ce domaine et présente les principaux résultats qui ont été démontrés ayant un lien avec cette thèse. Nous présentons les problématiques qui seront traitées dans les chapitres suivants, et expliquons notre contribution au domaine.

### **Théorie de la dualité et méthodes topologiques appliquées à la logique**

Les *dualités* permettent d'exprimer des relations entre deux différents phénomènes mathématiques qui peuvent être perçus, d'une certaine façon, comme équivalents. Le formalisme donné par la théorie des catégories nous donne la définition suivante : une dualité est une équivalence contravariante entre deux catégories. Le simple fait d'étudier une correspondance qui renverse le sens des flèches nous permet d'obtenir deux formulations différentes, et pourtant équivalentes, du même problème. Ce phénomène est répandu en mathématiques, et parfois, la compréhension d'un problème peut être grandement facilitée par l'étude de son analogue dual. On observe souvent des dualités entre les catégories de structures algébriques et les catégories de structures topologiques. En 1936, M. H. Stone a initié l'étude de la théorie de la dualité dans le cadre de la logique en présentant une dualité entre la catégorie des algèbres de Boole et la catégorie des espaces compacts Hausdorff possédant une base d'ouverts-fermés : les fameux *espaces Booléens* [67]. Cette dualité porte le nom de *dualité de Stone*

*Booléenne*. Celle-ci a d'abord été généralisée par Stone lui-même aux catégories de treillis distributifs bornés. Le pendant dual de cette catégorie est la catégorie dont les objets sont les *espaces spectraux* – c'est-à-dire, les espaces topologiques sobres, quasi-compacts, tels que l'intersection de deux ouverts quasi-compacts est quasi-compacte, et la collection des ouverts quasi-compacts forment une base pour la topologie; et dont les flèches sont les applications parfaites – c'est-à-dire, les applications continues telles que l'inverse d'un ouvert quasi-compact est un ouvert quasi-compact [68]. Plusieurs extensions de la dualité de Stone ont été étudiées depuis. Par exemple, Isbell et Papert [37] donnent une adjonction entre la catégorie dont les objets sont les espaces topologiques et dont les flèches sont les applications continues, et l'opposé de la catégorie dont les objets sont les frames et dont les flèches sont les morphismes de frames. Cette adjonction peut être restreinte afin d'obtenir une dualité entre les espaces sobres et les frames spatiales : voir [38] et [32] pour davantage de détails sur le contexte mathématique de la dualité de Stone et ses généralisations.



Plusieurs dualités de type Stone

La dualité de Stone et ses variantes sont fondamentales afin de faire le lien entre l'approche syntactique et l'approche sémantique de la logique. À l'origine, les algèbres de Boole ont été introduites comme un ensemble de règles formalisant le comportement de la logique classique. Cela a permis un traitement algébrique de cette dernière. La théorie de la dualité a ensuite permis de reformuler et repenser les propriétés qui apparaissent en logique en des termes purement topologiques. Ce genre de raisonnement ne se borne pas au cadre de la logique classique : de nombreux autres types de logiques admettent un traitement algébrique, ceci est au cœur de ce que l'on appelle la *logique algébrique abstraite* [26].

Logic	Algèbre	Espace topologique
Logique intuitioniste	Algèbre de Heyting	Espace d'Esakia
Logique modale S4	Algèbre intérieure	Espace Booléen partiellement ordonné
Logique classique	Algèbre de Boole	Espace Booléen
Logique pertinente	Algèbre pertinente	Espace d'Urquhart
Logique de Lukasiewicz	Algèbre MV	Espace de Tychonoff
$\mu$ -calcul modal	$\mu$ -algèbre modale	$\mu$ -frame modale
Logique de Markov	Algèbre d'Aumann	Processus de Stone Markov

Figure 1: Différentes logiques et leurs équivalents algébriques et topologiques, voir [24], [75], [46], [35] et [39]

### Classes de complexité et problème de séparation

La théorie de la complexité [49], [50], [2] s'attache à la classification des problèmes de calculs et s'intéresse aux liens existant entre les différentes classes ainsi obtenues. Cette théorie formalise l'intuition que l'on peut avoir de "l'efficacité de calcul" en introduisant des modèles mathématiques de la notion de calcul, et de la mesure de la quantité de ressources nécessaire à la résolution des problèmes, par exemple le temps et l'espace de stockage.

Un exemple d'un modèle de calcul abstrait sont les *circuits Booléens*, un modèle simplifié des circuits digitaux utilisés dans les ordinateurs modernes [77]. Les *classes de complexité* sont alors définies en considérant la taille de ces circuits, c'est-à-dire le nombre de sommets présents dans ceux-ci. Un autre exemple de modèle de calcul sont les *machines de Turing*, introduites en 1936 [74], qui sont des machines modélisant le traitement de l'information de manière très générale. La théorie de la complexité tente de déterminer les limites pratiques de ces différents modèles de calcul.

Le nom et la classe de complexité associée à chaque problème de décision	
<i>ALL</i>	tous les problèmes de décision
<i>NP</i>	décidable par une machine de Turing non déterministe en temps polynomial
<i>NL</i>	décidable par une machine de Turing non déterministe en temps logarithmique
<i>P</i>	décidable par une machine de Turing déterministe en temps polynomial
<i>L</i>	décidable par une machine de Turing déterministe en temps logarithmique
<i>ACC</i> <sup>0</sup>	décidable par des circuit booléens de profondeur constante, de taille polynomiale dont les portes sont ET, OU, NON, MODULAIRE, de degrés entrants non bornés
<i>AC</i> <sup>0</sup>	décidable par des circuit booléens de profondeur constante, de taille polynomiale, dont les portes sont ET, OU, NON, de degrés entrants non bornés

Quelques classes de complexité, voir [49] ou [50] pour d'autres exemples.

La théorie de la complexité ainsi définie est un sujet en apparence complètement inscrit dans l'informatique théorique. Toutefois, une approche plus mathématique est possible. La *théorie de la complexité descriptive* est une branche de la théorie de la complexité et de la théorie des modèles finis [43] dont le but est d'exprimer les classes de



complexité en utilisant des formules décrivant une logique sur certaines structures finies : *la logique sur les mots*. Nous présentons l'idée générale brièvement, un traitement plus détaillé sera apporté plus tard dans cette thèse. Pour tout alphabet fini  $A$ , nous pouvons voir un mot comme une structure

$$(\{0, \dots, |w| - 1\}, <, (a(\cdot))_{a \in A}),$$

où  $a(\cdot)$  est interprété comme l'ensemble des entiers naturels  $i$  tels que la  $i$ -ème lettre du mot  $w$  est un  $a$ , et  $<$  est l'ordre usuel sur les entiers. Par exemple, la proposition  $\exists x a(x)$  a pour interprétation "il existe une position  $x$  dans  $w$  telle que la lettre à la position  $x$  est un  $a$ ", ce qui correspond au langage  $A^*aA^*$ . Ce lien permet de traduire les résultats de la théorie de la complexité en terme de résultats de logique sur ces structures finies. Ceci permet de penser de nouvelles méthodes de preuves, et fournit un argument supplémentaire en la faveur de la "naturalité" des classes de complexité, dans le sens où elles ne sont pas purement attachées aux modèles de calculs utilisés pour les définir. En 1974, Fagin donne le premier résultat majeur en théorie de la complexité descriptive en prouvant que la classe de complexité  $NP$  peut être caractérisée comme la famille de langages correspondant aux propositions de la logique existentielle du second ordre [25]. Par la suite, Immerman prouve de nombreuses caractérisations de ce type pour d'autres classes de complexité, [36].

Les classes de complexité vues comme des fragments de logique	
$NP$	Logique existentielle du second ordre
$NL$	Logique du premier ordre avec un opérateur de clôture transitive
$P$	Logique du premier ordre avec un opérateur de plus petit point fixe
$L$	Logique du premier ordre avec un opérateur de clôture commutative et transitive
$ACC^0$	Logique du premier ordre avec quantificateur modulaire
$AC^0$	Logique du premier ordre

Fragments de logique associés à certaines classes de complexité, voir [36], [70].

Considérant deux classes de complexité  $\mathcal{C}_1$  et  $\mathcal{C}_2$ , le problème de séparation consiste à dire s'il existe un problème appartenant à  $\mathcal{C}_1$  mais pas à  $\mathcal{C}_2$ . Le problème de ce genre le plus connu est  $P \stackrel{?}{=} NP$ . La question est de savoir si les machines de Turing non déterministes polynomiales ont une puissance de calcul analogue aux machines de Turing déterministes polynomiales. Comme nous l'avons vu, beaucoup d'autres classes de complexité existent, et par conséquent, encore plus de problèmes de séparation existent encore. Toutefois, la plupart de ces questions restent sans réponse aujourd'hui encore. Notons, par exemple, que la classe  $ACC^0$  n'a été séparée d'aucune classe en dessous de  $NP$ .

Cela montre combien il reste à découvrir dans ce domaine. Obtenir de nouveaux résultats, même pour des classes de complexité correspondant à des puissances de

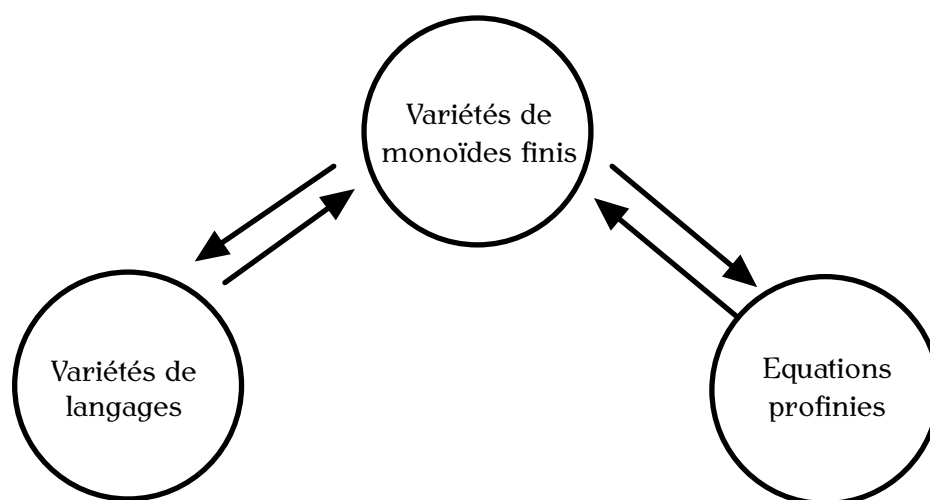
calcul moindres, pourrait mener à des perspectives intéressantes. Le peu de résultats de séparation qui ont été démontrés jusqu'à présent reposent essentiellement sur des méthodes purement combinatoires, algorithmiques et probabilistes, [80]. L'ingrédient que nous voulons ajouter provient d'un pan des mathématiques différent : il s'agit de la topologie.

## Reconnaissance de langages et équations topologiques

La théorie de la complexité descriptive nous permet de reformuler le problème de séparation de la manière suivante : soient deux classes de logique sur les mots  $\mathcal{L}_1$  et  $\mathcal{L}_2$ , peut-on trouver une formule dans  $\mathcal{L}_1$  qui n'est pas dans  $\mathcal{L}_2$  ? L'interprétation d'une formule dans une de ces classes de logique correspond à un langage défini sur un alphabet fini  $A$ . Ainsi, la question devient la suivante : est-il possible de trouver un langage  $L$  qui appartient à la famille de langages correspondant aux interprétations des formules dans la classe  $\mathcal{L}_1$ , mais qui n'appartient pas à la famille de langages correspondant aux interprétations des formules dans la classe  $\mathcal{L}_2$  ? Nous commençons par réduire notre attention au cas où tous les langages dans les familles que nous considérons sont réguliers.

Dans ce cas, la théorie des automates finis [63] ainsi que la théorie des monoïdes finis [51] nous fournissent des outils pour caractériser nos langages : il s'agit de la notion de *reconnaissance de langages*. Un langage  $L \subseteq A^*$  est reconnu par un monoïde  $M$  si il existe  $P \subseteq M$  et un morphisme de monoïde  $h : A^* \rightarrow M$  tel que  $h^{-1}(P) = L$ . Par extension, on dit aussi que  $M$  reconnaît  $L$ . À tout langage  $L \subseteq A^*$ , nous pouvons associer un monoïde fini  $M_L$ , que nous appelons le *monoïde syntactique* de  $L$ . Ce monoïde est le plus petit monoïde qui reconnaît  $L$  : cela signifie que  $M_L$  reconnaît  $L$ , et que pour tout monoïde  $M$  reconnaissant  $L$ ,  $M_L$  est un quotient d'un sous-monoïde de  $M$ . Nerode a prouvé qu'un langage est régulier si, et seulement si, son monoïde syntactique est fini [48]. Cela fait du monoïde syntactique un outil important pour la reconnaissance de langages, dans la mesure où il nous permet de remplacer un langage par un objet appartenant à la théorie, bien ancrée, des monoïdes finis. En 1965, Schützenberger prouve qu'un langage régulier est sans-étoile si, et seulement si, son monoïde syntactique est aperiodique, c'est-à-dire, si chaque élément  $x$  dans son monoïde syntactique est tel qu'il existe  $n \geq 0$  tel que  $x^n = x^{n+1}$  [64]. Ce résultat nous indique quelque chose de très important : il devrait y avoir une correspondance, non pas entre chaque monoïde fini et chaque langage régulier, mais plutôt entre certaines familles de monoïdes finis, et certaines familles de langages réguliers. En 1974, le théorème d'Eilenberg [23] donne un cadre général permettant d'appliquer la stratégie du résultat de Schützenberger en établissant une correspondance biunivoque entre les *variétés de langages réguliers* et les *variétés de monoïdes finis*, c'est-à-dire, les classes de monoïdes finis closes par sous-monoïde, quotient de monoïde et produit direct fini. Cette correspondance devient vraiment puissante une fois combinée avec une autre correspondance, démontrée par

Reiterman [61] en 1982. *Le théorème de Reiterman* est une variante du théorème des variétés de Birkhoff en algèbre universelle. Celui-ci dit que toute variété de monoïde fini est caractérisée par un ensemble d'*identités profinies*. Une identité profinie est une identité entre deux *mots profinis*. La définition précise de mot profini sera donnée dans le Chapitre 1. On peut voir ceux-ci comme des limites de suites de mots finis pour une certaine métrique, *la métrique profinie*. Par exemple, on peut montrer que, pour tout mot fini  $u$ , la suite  $(u^{n!})_{n \in \mathbb{N}}$  converge pour la métrique profinie : nous notons  $u^\omega$  sa limite.



### Correspondance d'Eilenberg-Reiterman dans le cas régulier

Les théorèmes d'Eilenberg et Reiterman ont été généralisés dans plusieurs directions durant ces vingt dernières années en assouplissant la notion de variété de langages, cf [40], [55] et [57]. L'outil permettant de généraliser ces résultats à n'importe quel algèbre de Boole de langages, réguliers ou non, est la dualité topologique. Celle-ci nous permet d'obtenir une correspondance biunivoque entre les familles de langages et leurs équations.

Ce lien avec la dualité est déjà présent dans le cadre régulier. Pippenger [58] est le premier à caractériser l'ensemble des mots profinis du point de vue de la dualité de Stone. Il a démontré que le dual de la sous-algèbre de Boole de  $\mathcal{P}(A^*)$  de tous les langages réguliers est l'espace des mots profinis, équipé avec la métrique profinie. Dans

[27], il a été démontré que la combinaison des correspondances Eilenberg-Reiterman est en fait un cas particulier de la correspondance plus générale donnée par la dualité de Stone entre sous-algèbre de Boole et espaces quotients. En effet, l'idée de considérer des équations dans le cadre régulier vient de ce que, si  $\mathcal{C}$  est une sous-algèbre de Boole de l'algèbre de Boole des langages réguliers sur un alphabet fini  $A$ , alors son dual est un quotient de l'espace dual à l'algèbre de Boole des langages réguliers, c'est-à-dire, l'ensemble des mots profinis sur  $A$ . Puisqu'il s'agit d'un quotient, cet espace est obtenu en égalisant des paires d'éléments de l'espace dual, autrement dit des mots profinis. Plus généralement, si  $\mathcal{B}$  est un algèbre de Boole de langages, réguliers ou non, il s'agit toujours d'une sous-algèbre de Boole de  $\mathcal{P}(A^*)$ . Le dual de cette dernière est la *compactification de Čech-Stone* de l'espace discret  $A^*$ , c'est-à-dire, l'ensemble des ultrafiltres de  $\mathcal{P}(A^*)$ , que nous notons  $\beta(A^*)$ .  $\mathcal{B}$  peut ainsi être caractérisée comme un ensemble de paires de points de l'espace dual  $\mathcal{S}(\mathcal{B})$ . Ces paires de points du dual sont les équations que l'on pourrait obtenir via la théorie de Eilenberg-Reiterman. Plus précisément, pour toute paire d'ultrafiltres  $\gamma_1, \gamma_2 \in \beta(A^*)$  une équation  $\gamma_1 \leftrightarrow \gamma_2$  est satisfaite pour un langage  $L \subseteq A^*$  si, et seulement si,

$$L \in \gamma_1 \iff L \in \gamma_2.$$

Nous parlons alors d'*équations ultrafiltres*.

**Theorem** ([65]). Toute algèbre de Boole de langages peut être définie par un ensemble d'équations ultrafiltres de la forme  $\gamma_1 \leftrightarrow \gamma_2$ , où  $\gamma_1, \gamma_2 \in \beta(A^*)$ .

Toutefois, la méthode permettant de spécifier de telles équations n'est pas claire, étant donné qu'il n'existe pas de preuve constructive de l'existence des ultrafiltres libres. Il n'existe pas de procédure qui permettrait de construire un ensemble d'équations "pratiques" directement à partir d'une algèbre de Boole de langages donnée : le problème d'appartenance n'est parfois pas décidable. Si l'algèbre de Boole que nous considérons est close par quotients, alors l'ensemble des équations qu'elle satisfait est un genre de congruence. Pour tout ultrafiltre sur les mots  $\gamma \in \beta(A^*)$  et pour tout mot  $w \in A^*$ , nous notons

$$w.\gamma := \{w^{-1}.L : L \in \gamma\} \text{ et } \gamma.w := \{L.w^{-1} : L \in \gamma\}.$$

Etant donnés  $\gamma_1, \gamma_2 \in \beta(A^*)$ , nous disons qu'un langage satisfait l'équation ultrafiltre  $\gamma_1 = \gamma_2$  s'il satisfait toutes les équations ultrafiltres  $w.\gamma_1 \leftrightarrow w.\gamma_2$  et  $\gamma_1.w \leftrightarrow \gamma_2.w$ , pour tout mot  $w \in A^*$ . Dans le cas particulier des algèbres de Booles de langages réguliers, un résultat encore plus fort existe.

**Theorem** ([29], Proposition 1.3). Soit  $\mathcal{B}$  une algèbre de Booles de langages réguliers sur  $A^*$  close par quotients, et soit  $w_1, w_2 \in \widehat{A^*}$ . Si  $\mathcal{B}$  satisfait l'équation profinie  $w_1 \leftrightarrow w_2$ , alors elle satisfait également les équations  $u.w_1 \leftrightarrow u.w_2$  et  $w_1.u \leftrightarrow w_2.u$ , pour tout mot profini  $u \in \widehat{A^*}$ .

Etant donné deux mots profinis  $w_1, w_2 \in \widehat{A^*}$ , on dit qu'un langage régulier  $L$  satisfait l'équation profinie  $w_1 = w_2$  s'il satisfait les équations profinies  $u.w_1 \leftrightarrow u.w_2$  et  $w_1.u \leftrightarrow w_2.u$ , pour tout mot profini  $u \in \widehat{A^*}$ . L'intérêt majeur de ces notations est de réduire le nombre d'équations nécessaires à la description d'une algèbre de Boole de langages qui est close par quotients.

### Fragment existentiel de la logique du premier ordre

Comme nous l'avons mentionné précédemment, la classe de circuits Booléens  $AC^0$  correspond au fragment de logique sur les mots correspondant aux formules du premier ordre écrites en utilisant des prédicats numériques arbitraires. L'ensemble des langages réguliers appartenant  $AC^0$  a été caractérisé par Barrington, Straubing et Therien [4]: ce sont exactement les langages satisfaisant les équations profinies

$$(u^{\omega-1}v)^{\omega+1} = (u^{\omega-1}v)^\omega,$$

pour tout  $u, v \in A^*$ . Ce résultat repose sur un autre résultat prouvé en 1984 par Furst [44], et qui n'admet actuellement aucune preuve purement algébrique. Plus généralement, aucune caractérisation topologique de l'algèbre de Boole de langages correspondant à  $AC^0$  n'est connue. L'étude de la dualité pour ce fragment est une tâche trop complexe pour être traitée dans son entièreté dans notre étude.

Dans cette thèse, nous nous concentrons sur un fragment significativement plus petit des formules du premier ordre : le fragment  $\mathcal{B}\Sigma_1[\mathcal{N}]$ , la clôture Booléenne de l'ensemble des propositions écrites en utilisant des prédicats numériques arbitraires, et sans alternance de quantificateurs. Aucune caractérisation de son espace dual n'est actuellement connue, et en particulier aucune caractérisation équationnelle n'est disponible. Certains résultats ont toutefois été établis pour des sous-fragments de  $\mathcal{B}\Sigma_1$ , nous en donnons quelques-uns dans les lignes suivantes. Le cadre régulier a été abondamment étudié, et plusieurs fragments intéressants de  $\mathcal{B}\Sigma_1$  ont été caractérisés par des équations profinies. Simon [66] a démontré que  $\mathcal{B}\Sigma_1[<]$ , l'algèbre de Boole des langages réguliers correspondant aux combinaisons Booléennes de propositions écrites en utilisant le prédicat binaire  $<$  et les prédicats sur les lettres est décrit par les équations profinies

$$(uv)^\omega = (vu)^\omega \text{ and } u^\omega = u^{\omega+1},$$

où  $u, v \in A^*$ . Cela correspond à la variété des monoïdes **J**-triviaux, que nous notons **J**. Un autre exemple est  $\mathcal{B}\Sigma_1[\mathcal{N}] \cap \text{Reg}$ , l'algèbre de Boole des langages réguliers correspondant aux combinaisons Booléennes de propositions écrites en utilisant les prédicats numériques arbitraires et les prédicats sur les lettres, qui correspond à la variété de monoïdes finis que nous notons **J \* LI \* MOD**, voir [15] et [45] pour davantage de détails.

Dans le cadre non régulier, des résultats sont encore disponibles à condition de re-

streindre notre attention aux formules construites en utilisant uniquement des prédicats numériques uniformes d'arité un. Dans [30], Gehrke, Petrisan et Reggio étudient les conséquences, du côté topologique, de la quantification existentielle sur une variable libre pour une formule donnée. Ils prouvent que, pour toute formule  $\varphi(x)$  avec une variable libre  $x$ , si l'on veut obtenir un reconnaisseur topologique de l'algèbre de Boole de langages modélisant la formule  $\exists\varphi(x)$ , il est suffisant de considérer l'espace  $\mathcal{V}(X) \times X$ , où  $X$  est un reconnaisseur topologique de l'algèbre de Boole de langages modélisant la formule  $\varphi(x)$  et  $\mathcal{V}$  est la construction d'un hyperespace de Vietoris. Toutefois, le reconnaisseur ainsi obtenu n'est, a priori, pas minimal : le reconnaisseur topologique syntactique, qui correspond à l'espace dual de l'algèbre de Boole en question, est un sous-espace de  $\mathcal{V}(X) \times X$ . Cette remarque motive le chapitre 2. Enfin, une contribution majeure de Gehrke, Grigorieff et Pin [29], motivant le chapitre 3, est la construction d'une famille d'équations ultrafiltres pour  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_1^u]$ , le fragment existentiel pour lequel nous restreignons notre attention aux prédicats numériques uniformes unaires et aux prédicats nullaires. Bien que ce soit encore très loin de prendre en compte tous les prédicats numériques, il s'agit du premier exemple d'une famille "concrète" d'équations ultrafiltres pour un fragment de logique défini par une famille de langages non réguliers. Notons qu'ils ont également réussi, en utilisant ce résultat, à retrouver les équations profinies, déjà connues, obtenues en intersectant cette algèbre de Boole de langages avec l'algèbre de Boole des langages réguliers.

## Aperçu de la thèse

Fixons un entier naturel  $k \geq 1$ , et considérons l'algèbre de Boole  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ . Celle-ci correspond au fragment de logique du premier ordre sur les mots constitués des combinaisons Booléennes de propositions définies en utilisant un bloc d'au plus  $k$  quantificateurs existentiels, les prédicats sur les lettres et les prédicats numériques uniformes d'arité  $l \in \{1, \dots, k\}$ . Dans cette thèse, notre contribution est une étude de cette algèbre de Boole du point de vue de la dualité topologique. Nous résumons les principaux résultats contenus dans ce manuscrit.

Le chapitre 1 est une introduction auto-contenue du matériel nécessaire à l'étude ultérieure de cette algèbre de Boole. Nous donnons une présentation de la théorie de la dualité topologique pour les algèbres de Boole, et nous détaillons le cas particulier de la dualité entre algèbre modale et hyperespace de Vietoris. Nous donnons ensuite une introduction à la reconnaissance des langages formels, que nous étendons au cadre topologique, puis nous donnons une introduction à la logique sur les mots. Finalement, nous terminons le chapitre en introduisant la notion d'équation ultrafiltre mentionnée dans cette introduction.

Le chapitre 2 s'intéresse à l'algèbre de Boole  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$  et son espace dual, que nous notons  $X_k$ . Notre contribution est une étude complète de cet espace dual, car il faut souligner qu'aucune caractérisation de celui-ci n'était connue en dehors du cas où  $k = 1$ .

Nous donnons plusieurs caractérisations de  $X_k$ . La première s'obtient en exploitant la dualité entre algèbres modales et hyperespaces de Vietoris, nous permettant d'identifier  $X_k$  comme un certain sous-espace de l'hyperespace de Vietoris sur  $\beta(\mathbb{N}^k)$  à la puissance  $A^k$ . La seconde s'obtient via une approche consistant à considérer les éléments de  $X_k$  comme des "mots généralisés". Nous fixons un coloriage fini de  $\mathbb{N}^k$ , et nous expliquons comment il est parfois possible de définir un véritable mot fini correspondant à un élément de  $\mathcal{V}(\beta(\mathbb{N}^k))^{A^k}$ , du point de vue de ce coloriage. Nous prouvons alors que  $X_k$  consiste exactement en l'ensemble des points de  $\mathcal{V}(\beta(\mathbb{N}^k))^{A^k}$  qui vérifient cette propriété pour chaque coloriage fini de  $\mathbb{N}^k$ .

Le chapitre 3 traite la question des équations ultrafiltres pour  $\mathcal{B}\Sigma_1[\mathcal{N}_1^u]$ . Une famille d'équations ultrafiltres pour  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_1^u]$ , le fragment obtenu à partir de  $\mathcal{B}\Sigma_1[\mathcal{N}_1^u]$  en ajoutant les prédicats numériques nullaires, a déjà été introduite dans [29]. Notre première contribution à ce sujet est une présentation plus topologique de ces équations ultrafiltres. Notre approche est fondée sur l'idée qu'il est possible de reformuler les équations en question via une condition sur les coloriages finis de  $\mathbb{N}$ . De cette manière, nous arrivons à grandement réduire la quantité de raisonnement combinatoires requis dans [29] pour prouver la correction et la complétude de ces équations. Notre deuxième contribution est l'utilisation de cette approche pour trouver une base d'équations ultrafiltres pour  $\mathcal{B}\Sigma_1[\mathcal{N}_1^u]$ , et la vérification de leur correction et complétude. Nous concluons la thèse par quelques pistes permettant de poursuivre l'étude pour  $k \geq 1$ .

This thesis fits in the area of research that investigates the application of topological duality methods to problems that appear in theoretical computer science. More specifically, our eventual goal is to derive results in computational complexity theory by studying appropriate topological objects which characterise them. The link which relates these two seemingly separated fields is logic, more precisely a subdomain of finite model theory known as *logic on words*. It allows for a description of complexity classes as certain families of languages, possibly non-regular, on a finite alphabet.

This introduction gives an historical account on the problem, and explains the progress that have been made prior to this thesis. It introduces the problematic that will be treated in the next chapters, and explains our contribution to the domain.

### Duality theory and topological methods in logic

*Dualities* are a way to express the relationship between two different mathematical phenomenon which can be seen as equivalent, in a certain sense. Through the formalism of category theory, a straight-forward definition is the following: a duality is a contravariant equivalence between two categories. The simple fact of looking at a correspondence which reverses arrows is the perspective that allows us to obtain two different, yet equivalent, formulations of the same problem. Dualities are widespread in mathematics, and depending on the problem, one might benefit from an easier understanding by looking at things on the dual side. Often, a duality is between a category of algebraic structures and a category of topological structures. In 1936, M. H. Stone initiated duality theory in logic by presenting a duality between the category of Boolean algebras and the category of compact Hausdorff spaces having a basis of clopen sets, so-called *Boolean spaces* [67]. This is referred to as *Boolean Stone duality*. A first generalization has been done by Stone himself, extending his duality to the category of bounded distributive lattices and lattice homomorphisms. The dual counterpart is the category whose objects are *spectral spaces*, that is topological spaces that are sober,



quasi-compact, such that the intersection of two quasi-compact opens is quasi-compact, and the collection of quasi-compact opens forms a basis for the topology, and whose arrows are perfect maps, that is continuous maps such that the inverse image of a quasi-compact open is quasi-compact [68]. Many other extensions of Stone duality have been studied since then. For instance, Isbell and Papert [37] give an adjunction between the category of topological spaces with continuous functions and the opposite category of frames with frame homomorphisms. This restricts to a duality between sober spaces and spatial frames: see [38] and [32] for more details on the mathematical context of Stone duality and its generalizations.

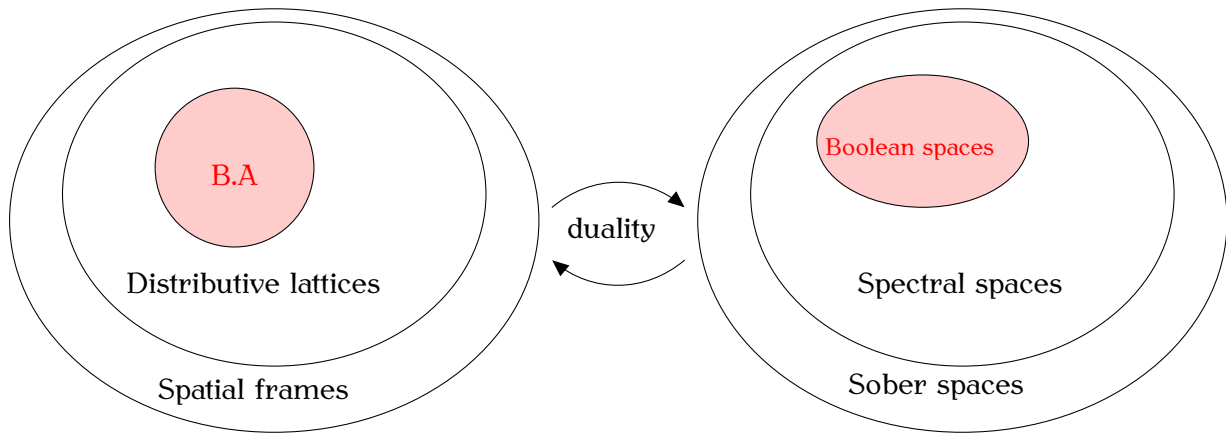


Figure 2: Some Stone type dualities

Stone's duality and its variants are central in making the link between syntactical and semantic approaches to logic. Originally, Boolean algebras were introduced as a set of rules which formalized the behaviour of classical logic. This allowed for an algebraic treatment of classical logic. Duality theory then enabled the rephrasing of properties that appear in logic in order to think about them in purely topological terms. This kind of reasoning does not only apply to classical logic: many other logics admit an algebraic treatment, this is the heart of *abstract algebraic logic* [26].

Logic	Algebra	Topological space
Intuitionistic logic	Heyting algebra	Esakia space
S4 modal logic	Interior algebra	Partially ordered Boolean space
Classical logic	Boolean algebra	Boolean space
Relevant logic	Relevant algebra	Urquhart space
Lukasiewicz logic	MV algebra	Tychonoff space
Modal $\mu$ -calculus	Modal $\mu$ -algebra	Modal $\mu$ -frame
Markovan logic	Aumann algebra	Stone Markov process

Figure 3: Different logics and their algebraic and topological counterparts, see [24], [75], [46], [35] and [39]

### Complexity classes and separation problem

Computational complexity theory [49], [50], [2] focuses on putting computational problems in different classes and relating these classes to each other. The theory formalizes the intuition we have on “computation efficiency” by introducing mathematical models of computation and measuring the amount of resources they require to solve problems, for instance time and storage. An example of such models are *Boolean circuits*, a simplified model of the digital circuits used in modern computers [77]. *Circuit complexity classes* are defined in terms of circuit size, that is the number of vertices in the circuit. Another example are *Turing machines*, introduced in 1936 [74], which are machines modelling information processing in a very general fashion. Computational complexity theory attempts to determine the practical limits of these computational models.

The name and the corresponding class of decision problems	
<i>ALL</i>	all decision problems
<i>NP</i>	solvable by a non-deterministic Turing machine in polynomial time
<i>NL</i>	solvable by a non-deterministic Turing machine in logarithmic time
<i>P</i>	solvable by a deterministic Turing machine in polynomial time
<i>L</i>	solvable by a deterministic Turing machine in logarithmic time
<i>ACC</i> <sup>0</sup>	solvable by a family of constant depth unlimited-fanin AND, OR, NOT, MODULAR gates, polynomial-size circuits
<i>AC</i> <sup>0</sup>	solvable by a family of constant depth unlimited-fanin AND, OR, NOT gates, polynomial-size circuits

Figure 4: A few computational complexity classes, for more see [49] or [50].

Complexity theory defined as such is a topic that seems to purely belong to theoretical computer science. However, there exists a way to think about it in more mathematical terms. *Descriptive complexity* is a branch of computational complexity theory and of finite model theory [43] whose purpose is to express complexity classes in terms of formulas for a logic on a finite structure: *logic on words*. We briefly expose it here, but explain it more in detail in the thesis. For any finite alphabet  $A$ , we view a word as a structure

$$(\{0, \dots, |w| - 1\}, <, (a(\cdot))_{a \in A}),$$

where  $a(\cdot)$  is interpreted as the set of integers  $i$  such that the  $i$ -th letter of  $w$  is an  $a$ , and  $<$  is the usual order on integers. For instance, the sentence  $\exists x a(x)$  is interpreted as “there is a position  $x$  in  $w$  such that the letter in position  $x$  is  $a$ ”, which corresponds to the language  $A^*aA^*$ . This connection enables to translate results of complexity theory into results on the logic of finite structures, facilitating new proof methods and providing additional evidence that the main complexity classes are somehow “natural” and not tied to the specific abstract machines used to define them. In 1974, Fagin provided the first major result in descriptive complexity theory by proving that the complexity class  $NP$  is characterised as the family of languages corresponding to sentences of existential second-order logic [25]. Following this result, Immerman proved numerous characterisations of this kind for other complexity classes [36].

Complexity classes as logic fragments	
$NP$	Existential second-order logic
$NL$	First-order logic with a transitive closure operator
$P$	First-order logic with a least fixed point operator
$L$	First-order logic with a commutative, transitive closure operator
$ACC^0$	First-order logic with modular quantifier
$AC^0$	First-order logic

Figure 5: Logic fragments associated to some computational complexity classes, see [36], [70]

Considering two complexity classes  $\mathcal{C}_1$  and  $\mathcal{C}_2$ , the separation problem in complexity theory consists in telling whether there exists a problem that belong to  $\mathcal{C}_1$  and not to  $\mathcal{C}_2$ . The most well-known problem of this kind is known as  $P \stackrel{?}{=} NP$ . It asks whether the non-deterministic polynomial time Turing machines have more computational power than deterministic polynomial time Turing machines. There are plenty of other complexity classes, and the question of separation for most of them remains unanswered. Note, for instance, that the class  $ACC^0$  has not been separated from anything all the way up to  $NP$ .

This shows how much there is to discover in this field, and enforces the fact that obtaining results even for classes with a weaker computational power such as Boolean circuits would potentially lead to interesting perspectives. The few separation results that have been proved rely on using combinatorial and probabilistic, as well as algorithmic methods [80]. The ingredient we want to develop comes from a different area of mathematics: topology.

## Language recognition and topological equations

Descriptive complexity theory allows to rephrase the separation problem as asking, for two fixed classes of logic on words  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , whether it is possible to find a formula in  $\mathcal{L}_1$  which does not belong to  $\mathcal{L}_2$ . The interpretation of a formula in a logic class corresponds to a language defined on a finite alphabet  $A$ . Therefore, the question now becomes whether it is possible to find a language  $L$  which belongs to the family of languages corresponding to the interpretations of the formulas in the class  $\mathcal{L}_1$ , but not to the family of languages corresponding to the interpretations of the formulas in the class  $\mathcal{L}_2$ . Let us start by restricting our attention to the case where the families of languages we consider only involve regular languages.

In that situation, we find tools to characterise our languages in finite automata theory [63] and finite monoid theory [51]: this is the notion of *language recognition*. A language  $L \subseteq A^*$  is recognized by a monoid  $M$  if there exists  $P \subseteq M$  and a monoid morphism  $h : A^* \rightarrow M$  such that  $h^{-1}(P) = L$ . By extension, we also say that  $M$  recognizes  $L$ . To every language  $L \subseteq A^*$ , we can associate a monoid  $M_L$ , that we refer to as the *syntactic monoid* of  $L$ . This monoid is the smallest monoid that recognizes  $L$ : that is,  $M_L$  recognizes  $L$ , and for every monoid  $M$  that recognizes  $L$ ,  $M_L$  is a quotient of a submonoid of  $M$ . Nerode proved that a language is regular if, and only if, its syntactic monoid is finite [48]. This makes the syntactic monoid an important tool in studying recognisable languages, since it allows us to replace the language by an object of the well-rooted theory of finite monoids. In 1965, Schützenberger proves that a regular language is star-free if, and only if, its syntactic monoid is aperiodic, that is, if every element  $x$  in the syntactic monoid is such that there exists  $n \geq 0$  such that  $x^n = x^{n+1}$  [64]. This results point us toward something important: there should be a correspondence, not between individual finite monoids and individual recognisable languages, but between families of finite monoids and families of regular languages. In 1974, Eilenberg’s theorem [23] supplied a general framework in which to apply the strategy of Schützenberger’s result by stating that *varieties of regular languages* are in one-to-one correspondence with *varieties of finite monoids*, that is, classes of finite monoids closed under taking submonoids, quotient monoids and finite direct products. This correspondence becomes very powerful once it has been combined with another correspondence proved by Reiterman [61] in 1982. Reiterman’s theorem is a variant of Birkhoff’s variety theorem from universal algebra. It states that any variety of finite

monoids can be characterized by a set of *profinite identities*. A profinite identity is an identity between two *profinite words*. The precise definition of profinite words will be given in Chapter 1. They can be viewed as limits of sequences of finite words for a certain metric, *the profinite metric*. For instance, one can show that, for every finite word  $u$ , the sequence  $(u^{n!})_{n \in \mathbb{N}}$  converges for the profinite metric: we denote by  $u^\omega$  its limit.

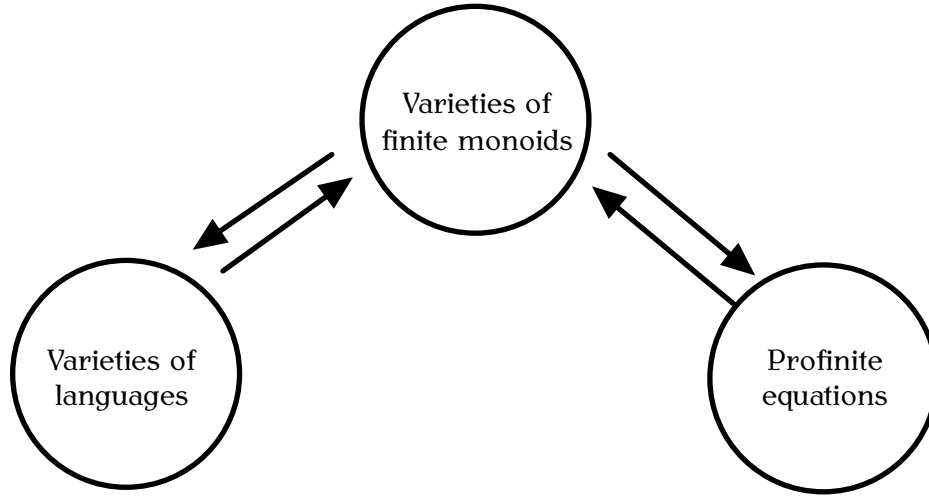


Figure 6: Eilenberg-Reiterman correspondence in the regular case

Eilenberg's and Reiterman's theorems have been extended several times over the last twenty years by relaxing the definition of a variety of languages, cf [40], [55] and [57]. The tool that allows for a generalization of this result to any Boolean algebra of languages, regular or not, is duality theory. It allows for a one-to-one correspondence between families of languages and their equations. This link with duality is already present in the regular setting. Pippenger [58] was the first to characterize the set of profinite words under the light of Stone duality. He proved that the dual of the sub-boolean algebra of  $\mathcal{P}(A^*)$  consisting in all regular languages is the space of all profinite words, equipped with the profinite metric. In [27], it was shown that the Eilenberg-Reiterman combination is in fact a special instance of the Stone duality between sub-Boolean algebras and quotient spaces. Indeed, the general mechanism of equations in the theory of regular languages arises from the fact that, if  $\mathcal{C}$  is a sub-Boolean algebra

of the Boolean algebra of regular languages on a given alphabet  $A$ , then its dual space is a quotient of the dual of the Boolean algebra of regular languages, that is, the set of all profinite words on  $A$ . As a quotient, it is given by equating elements in this dual space, which are profinite words. Now, more generally, if  $\mathcal{B}$  is a Boolean algebra of not necessarily regular languages, it is also a subalgebra of  $\mathcal{P}(A^*)$ . The dual of the latter is the *Čech-Stone compactification* of the discrete space  $A^*$ , that is, the set of all ultrafilters of  $\mathcal{P}(A^*)$ , which we denote by  $\beta(A^*)$ . It follows that  $\mathcal{B}$  may be characterised by a set of pairs of points of the dual space  $\mathcal{S}(\mathcal{B})$ . These pairs of points of the dual space are the equations of Eilenberg-Reiterman theory. More precisely, for any two ultrafilters  $\gamma_1, \gamma_2 \in \beta(A^*)$  an equation  $\gamma_1 \leftrightarrow \gamma_2$  holds of a language  $L \subseteq A^*$  if, and only if,

$$L \in \gamma_1 \iff L \in \gamma_2.$$

We refer to these as *ultrafilter equations*.

**Theorem** ([65]). Any Boolean algebra of languages can be defined by a set of ultrafilter equations of the form  $\gamma_1 \leftrightarrow \gamma_2$ , where  $\gamma_1, \gamma_2 \in \beta(A^*)$ .

However, it is not clear how to specify such equations since there is no constructive proof of the existence of free ultrafilters. There exists no procedure that would allow for a way to build practical equational basis out of a given Boolean algebra of languages: membership is sometimes not decidable. If the Boolean algebra we consider is closed under quotients, then the set of all equations satisfied by it is a kind of congruence. For any ultrafilter on words  $\gamma \in \beta(A^*)$  and any word  $w \in A^*$ , set

$$w.\gamma := \{w^{-1}.L : L \in \gamma\} \text{ and } \gamma.w := \{L.w^{-1} : L \in \gamma\}.$$

Given  $\gamma_1, \gamma_2 \in \beta(A^*)$ , we say that a language satisfies the ultrafilter equation  $\gamma_1 = \gamma_2$  if it satisfies all the ultrafilter equations  $w.\gamma_1 \leftrightarrow w.\gamma_2$  and  $\gamma_1.w \leftrightarrow \gamma_2.w$ , for each word  $w \in A^*$ . In the particular case of Boolean algebra of regular languages, an even stronger result is available.

**Theorem** ([29], Proposition 1.3). Let  $\mathcal{B}$  be a Boolean algebra of regular languages of  $A^*$  closed under quotients and let  $w_1, w_2 \in \widehat{A^*}$ . If  $\mathcal{B}$  satisfies the profinite equation  $w_1 \leftrightarrow w_2$ , then it satisfies the profinite equations  $u.w_1 \leftrightarrow u.w_2$  and  $u.w_1 \leftrightarrow w_2.u$ , for each profinite word  $u \in \widehat{A^*}$ .

Given  $w_1, w_2 \in \widehat{A^*}$ , we say that a regular language satisfies the profinite equation  $w_1 = w_2$  if it satisfies the profinite equations  $u.w_1 \leftrightarrow u.w_2$  and  $u.w_1 \leftrightarrow w_2.u$ , for each profinite word  $u \in \widehat{A^*}$ . The main interest of these notations is to allow one to produce smaller sets of defining equations for Boolean algebra of languages which are closed under quotients.

## Existential fragment of first-order logic

As we mentioned previously, the class of Boolean circuits  $AC^0$  corresponds to the fragment of logic on words corresponding to first-order formulas, with arbitrary numerical predicates. The set of regular languages which belong to  $AC^0$  has been characterised by Barrington, Straubing and Therien [4]: they are exactly the languages satisfying the profinite equations

$$(u^{\omega-1}v)^{\omega+1} = (u^{\omega-1}v)^\omega,$$

for any  $u, v \in A^*$ . This result relies on another result proved in 1984 by Furst [44], for which no purely algebraic proof is known. More generally, no topological characterisation of the Boolean algebra of languages corresponding to  $AC^0$  is known, and providing a duality theoretic treatment for this whole fragment is a task that is too complex to be conducted in our study.

In this thesis, we focus on a significantly smaller fragment of first-order formulas: the fragment  $\mathcal{BS}_1$  which is the Boolean closure of the set of sentences written with arbitrary numerical predicates and without any quantifier alternation. No characterisation of its dual space is currently known, not even mentioning an equational characterisation. Some results have however been established for some smaller fragments of  $\mathcal{BS}_1[\mathcal{N}]$  that we summarize now. The regular setting has been abundantly studied, and several interesting fragments of  $\mathcal{BS}_1$  have been given a characterisation in terms of profinite equations. It has been proven by Simon [66] that  $\mathcal{BS}_1[<]$ , the Boolean algebra of regular languages corresponding to Boolean combinations of sentences written by using the binary numerical predicate  $<$  and letter predicates is described by the profinite equations

$$(uv)^\omega = (vu)^\omega \text{ and } u^\omega = u^{\omega+1},$$

where  $u, v \in A^*$ . This corresponds to the variety of **J**-trivial monoids, denoted by **J**. Another example is  $\mathcal{BS}_1[\mathcal{N}] \cap \text{Reg}$ , the Boolean algebra of regular languages corresponding to Boolean combinations of sentences written by using any numerical predicate and letter predicates, which has been proven to correspond to the variety of finite monoids denoted by **J** \* **LI** \* **MOD**, see [15] and [45] for more details.

In the non-regular setting, some results are also available, as long as we restrict our attention to formulas built by using only uniform numerical predicates of arity one. In [30], Gehrke, Petrisan and Reggio study the topological construction corresponding to adding a layer of existential quantification over one free variable in a formula. They prove that, for any formula  $\varphi(x)$  with a free variable  $x$ , in order to obtain a topological recognizer for the Boolean algebra of languages which models the formula  $\exists \varphi(x)$ , it is enough to consider the space  $\mathcal{V}(X) \times X$ , where  $X$  is a topological recognizer of the Boolean algebra of languages which models the formula  $\varphi(x)$  and  $\mathcal{V}$  is the Vietoris hyperspace construction. However, this recognizer is a priori not minimal: the syntactic recognizer, which corresponds to the dual space, is a subspace of  $\mathcal{V}(X) \times X$ . This remark

is what motivates Chapter 2. Finally, a major contribution of Gehrke, Grigorieff and Pin [29], which motivates Chapter 3, is the construction of a family of ultrafilter equations for  $\mathcal{BS}_1[\mathcal{N}_0, \mathcal{N}_1^u]$ , the existential fragment where we restrict our attention to only nullary numerical predicates and unary uniform numerical predicates. Even though this is still far from taking into account every numerical predicate, this is the first example of a “concrete” family of ultrafilter equations for a fragment defined by using non-regular languages. Note that they were also able to retrieve the profinite equations that had already been established for the restriction of this fragment to regular languages.

## Overview of the thesis

Fixing an integer  $k \geq 1$ , we consider the Boolean algebra  $\mathcal{BS}_1[\mathcal{N}_k^u]$ . It corresponds to the fragment of first-order logic on words consisting in Boolean combinations of sentences defined by using a block of at most  $k$  existential quantifiers, letter predicates and uniform numerical predicates of arity  $l \in \{1, \dots, k\}$ . Our contribution in this thesis is a duality-theoretic study of this Boolean algebra. We summarize the main results presented in this manuscript.

Chapter 1 is a self-contained introduction to all of the material required in order to conduct this study. We give a presentation of duality theory for Boolean algebras, and we detail the particular case of the relationships between modal algebras and Vietoris hyperspaces. We then present the notion of formal language recognition, that we extend to the topological setting, and we give an introduction to logic on words. We conclude the chapter by giving more details on the notion of ultrafilter equations that we mentioned in this introduction. Chapter 2 is concerned with the Boolean algebra  $\mathcal{BS}_1[\mathcal{N}_k^u]$  and its dual space, that we denote by  $X_k$ . Our contribution is a complete study of this dual space, as it should be noted that a characterisation of  $X_k$  was only known in the case  $k = 1$ . We provide several characterisations of  $X_k$ . First, by exploiting the duality between modal algebras and Vietoris hyperspaces, we identify  $X_k$  as a certain subspace of the Vietoris hyperspaces on  $\beta(\mathbb{N}^k)$  to the power  $A^k$ . Second, we follow an approach that, broadly speaking, relies on us viewing the elements of  $X_k$  as “generalized words”. Fixing a finite colouring of  $\mathbb{N}^k$ , we explain how it is sometimes possible to define an actual finite word which corresponds to an element in  $\mathcal{V}(\beta(\mathbb{N}^k))^{A^k}$ , with respect to this finite colouring. We prove that  $X_k$  consists exactly in the elements of  $\mathcal{V}(\beta(\mathbb{N}^k))^{A^k}$  that verify this property for every single finite colouring of  $\mathbb{N}^k$ . Chapter 3 treats the question of ultrafilter equations for  $\mathcal{BS}_1[\mathcal{N}_1^u]$ . A family of ultrafilter equations for  $\mathcal{BS}_1[\mathcal{N}_0, \mathcal{N}_1^u]$ , the fragment obtained from  $\mathcal{BS}_1[\mathcal{N}_1^u]$  by adding nullary numerical predicates, has already been introduced in [29]. Our first contribution is a more topological presentation of these ultrafilter equations. Our approach is based on the idea that it is possible to reformulate the ultrafilter equations in question in terms of a certain condition over finite colourings of  $\mathbb{N}$ . By doing so, we greatly reduce the amount of combinatorics required in [29] to prove the soundness and completeness of these equation. Our second



contribution is to use this approach to find a basis of ultrafilter equations for  $\mathcal{B}\Sigma_1[\mathcal{N}_1^u]$ , and check their soundness and completeness. We conclude the thesis with some hints in order to pursue the study for  $k \geq 1$ .

---

## Duality and formal language theory

---

This chapter consists in an introduction to the notions of duality theory and formal language theory necessary in order to understand the upcoming chapters. While the following sections are not written as a thorough introduction to the domains they treat, the notions are presented in a way that is coherent with our goals, and makes this document self-contained. In particular, most of the proofs are sketched, and we provide references whenever necessary.

**Outline of the chapter:** In Section 1.1, we recall definitions for lattice-like structures and we introduce Stone duality for Boolean algebras, which is the main tool we use all along the thesis. In Section 1.2, we detail the particular case of the duality between modal algebras and Vietoris hyperspaces, extensively used in Chapter 2, and provide a few instances of the correspondence between closed subsets of the dual space of a given Boolean algebra, and the space of filters of this Boolean algebra. In Section 1.3, we recall the basics on language theory, and introduce language recognition by finite monoids, which we generalize to the topological setting. In Section 1.4, we present the link that has been made between language theory and logic by giving an overview of logic on words. Finally, in Section 1.5, we introduce ultrafilter equations of a Boolean algebra. For any finite alphabet  $A$ , these allow in particular for a description of Boolean subalgebras of  $\mathcal{P}(A^*)$  in terms of a family of pairs of points in  $\beta(A^*)$ , and thus can be used in order to prove separation results for some fragment of logic on words.

### 1.1 Boolean algebras and Stone duality

In this section, we detail the contravariant category equivalence which exists between the category of Boolean algebras, with Boolean algebra homomorphisms, and a sub-

category of the category of topological spaces, which we refer to as the category of Boolean spaces. We start by explaining how this correspondence works in the case where the Boolean algebra we consider is finite. We then generalize it to arbitrary Boolean algebras, and provide a few examples of this correspondence which will be useful in the next chapters. The readers interested in a deeper treatment of lattice-like structures can refer to [20] and [13]. For duality theory, they can refer to [67] and [38].

### 1.1.1 General duality theory

The theory of lattices lies at the intersection of order theory and universal algebra. From the point of view of order theory, a *lattice* is a partially ordered set  $(L, \leq)$  such that, for any  $b, b' \in L$ , the *supremum*, which we denote by  $b \vee b'$  and the *infimum*, which we denote by  $b \wedge b'$ , exist. An example of lattice is given by the set of natural numbers, partially ordered by divisibility, for which the supremum is the least common multiple, and the infimum is the greatest common divisor.

A lattice  $(L, \leq)$  is said to be *distributive* if  $\vee$  distributes over  $\wedge$  and vice versa. It is said to be *bounded* if it contains a top element 1 and a bottom element 0. In particular, two elements in a bounded lattice such that their infimum is 0 are said to be *disjoint*. Finally, for any  $b \in L$ , a *complement* of  $b$  is an element  $c$  of  $L$  such that

$$b \wedge c = 0 \text{ and } b \vee c = 1.$$

In the particular case of distributive lattices, a complement of an element, if it exists, is then unique. We denote it by  $\neg b$ .

**Definition 1.1.** [Boolean algebras] A *Boolean algebra* is a bounded distributive lattice such that every element admits a complement. For any Boolean algebras  $\mathcal{B}_1, \mathcal{B}_2$ , a Boolean algebra homomorphism is a map  $f : \mathcal{B}_1 \rightarrow \mathcal{B}_2$  such that, for any  $b, b' \in \mathcal{B}_1$ ,  $f(b \wedge_{\mathcal{B}_1} b') = f(b) \wedge_{\mathcal{B}_2} f(b')$ ,  $f(b \vee_{\mathcal{B}_1} b') = f(b) \vee_{\mathcal{B}_2} f(b')$  and  $f(\neg b_1) = \neg f(b_1)$ . In particular, we have that  $f(0_{\mathcal{B}_1}) = 0_{\mathcal{B}_2}$  and  $f(1_{\mathcal{B}_1}) = 1_{\mathcal{B}_2}$ .

We denote by **Bool** the category for which the objects are Boolean algebras and morphisms are Boolean algebra homomorphisms.

The main example of a Boolean algebra is the powerset algebra of a set  $S$ ,  $\mathcal{P}(S)$ , considered with set-theoretic union, intersection, and complement. An important observation is the following: in order to reconstruct the Boolean algebra  $\mathcal{P}(S)$ , the data of the singletons  $\{s\}$ , for every  $s \in S$ , is actually sufficient. In terms of the order, singletons can be characterized as atoms.

**Definition 1.2.** An *atom* of a Boolean algebra  $\mathcal{B}$  is an element which is minimal among non-bottom elements of  $\mathcal{B}$ .

In particular, any two atoms are disjoint. Every finite Boolean algebra  $\mathcal{B}$  possesses atoms. In the particular case of the powerset algebra  $\mathcal{P}(S)$  of a finite set  $S$ , the atoms

are indeed the singletons  $\{s\}$ , for every  $s \in S$ , and they allow us to recover the full Boolean algebra.

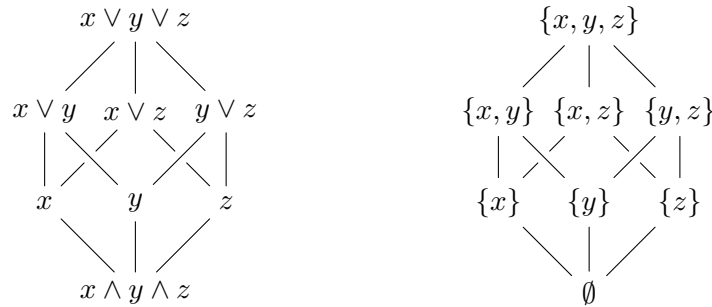


Figure 7: The powerset algebra on  $\{x, y, z\}$ , where  $x, y$  and  $z$  are pairwise disjoint elements.

This situation remains similar in the context of any finite Boolean algebra.

**Proposition 1.3.** *Any finite Boolean algebra is isomorphic to the powerset algebra on the set of its atoms.*

*Proof.* This result is folklore. For a proof, see, e.g., [20], Theorem 5.5.  $\square$

This reasoning still holds for certain infinite Boolean algebras such as powerset algebras on an infinite set  $S$ . However, it does not hold for an arbitrary Boolean algebra. Anticipating on Section 1.3, let us provide a simple example of an infinite Boolean algebra that is not of the form  $\mathcal{P}(S)$ , for any set  $S$ , stemming from language theory. If  $A$  is a finite alphabet, and  $\text{Reg}(A^*)$  is the Boolean algebra of all regular languages on  $A^*$ , then the atoms of this Boolean algebra are the singletons  $\{w\}$ , for any  $w \in A^*$ . The powerset of the set of atoms of  $\text{Reg}(A^*)$  is  $\mathcal{P}(A^*)$ , and these two Boolean algebras are not isomorphic, as there are countably many regular languages (every regular language corresponds to a finite automaton), but uncountably many languages. Even worse, one can construct infinite Boolean algebras which have no atom at all (cf [33], Chapter 16), such as for example the Boolean algebra of clopen sets of the Cantor space. The major insight of Stone was that, despite all of these facts, it is possible to generalize the duality introduced in the finitary case to arbitrary Boolean algebras. In order to do so, we have to take into account “generalized elements” which are incarnated by filters. This leads us to introducing a notion which generalizes the one of atom to the infinite case: the so-called ultrafilters.

**Definition 1.4.** Fix  $\mathcal{B}$  a Boolean algebra. A *filter*  $\mathcal{F}$  of  $\mathcal{B}$  is a non-empty subset of  $\mathcal{B}$  satisfying the following properties.

- For every  $b$  in  $\mathcal{F}$ , and every  $b'$  in  $\mathcal{B}$  such that  $b \leq b'$ , we have that  $b'$  is in  $\mathcal{F}$ .
- For every  $b$  and  $b'$  in  $\mathcal{F}$ ,  $b \wedge b'$  is in  $\mathcal{F}$ .

If  $\mathcal{F}$  is a filter that is not equal to  $\mathcal{B}$ , we say that  $\mathcal{F}$  is *proper*. A *filter basis* of a filter  $\mathcal{F}$  is a non-empty family  $B \subseteq \mathcal{F}$  such that, for all  $b' \in \mathcal{F}$ , there exists  $b \in B$  such that  $b \leq b'$ .

We denote by  $Filt(\mathcal{B})$  the set of filters of  $\mathcal{B}$ . For any set  $S$ , we use notation  $Filt(S)$  to refer to the set of filters of the powerset algebra  $\mathcal{P}(S)$ , and we abusively refer to those as the filters on  $S$ .

**Example 1.5.** We provide the two following examples of filters, which will be often used throughout the thesis.

- For any Boolean algebra  $\mathcal{B}$ , and for any element  $b \in \mathcal{B}$ , the set

$$\uparrow b := \{b' \in \mathcal{B} : b \leq b'\}$$

is a filter in  $\mathcal{B}$  that we refer to as the *principal filter* containing  $b$ .

- For any set  $S$ , the set of all cofinite subsets of  $S$

$$Cof(S) := \{T \subseteq S : T^c \text{ is finite}\}$$

is a filter that we refer to as the *Fréchet filter*.

The fundamental idea behind Stone's construction is that, by equipping the set of all ultrafilters of  $\mathcal{B}$  with an appropriate topology, it is possible to recover the Boolean algebra  $\mathcal{B}$ .

**Definition 1.6.** An *ultrafilter*  $\gamma$  of  $\mathcal{B}$  is a proper filter of  $\mathcal{B}$  such that, for every  $b$  in  $\mathcal{B}$ ,  $b$  or  $\neg b$  is in  $\gamma$ . We denote by  $\mathcal{S}(\mathcal{B})$  the space of all ultrafilters of  $\mathcal{B}$ , endowed with the topology generated by the sets of the form

$$\hat{b} := \{\gamma \in \mathcal{S}(\mathcal{B}) : b \in \gamma\},$$

for every  $b$  in  $\mathcal{B}$ . We refer to  $\mathcal{S}(\mathcal{B})$  as the *dual space* of  $\mathcal{B}$ .

*Remark 1.7.* We defined filters and ultrafilters of Boolean algebras since they are our main object of study in this thesis, however these notions still make sense in some similar frameworks. The definition of filter we gave only requires a semi-lattice structure in order to hold, and we can define ultrafilters of a lattice as filters that are maximal for inclusion among proper filters. In the case where the lattice is a Boolean algebra, this is equivalent to Definition 1.6. In the Boolean case, it is also equivalent to saying that  $\gamma$  is a *prime filter*: a proper filter  $\gamma$  such that, for every  $b, b' \in \mathcal{B}$  with  $b \vee b' \in \gamma$ , we have  $b \in \gamma$  or  $b' \in \gamma$ .

**Example 1.8.** We review Example 1.5 under the light of ultrafilters.

- For any Boolean algebra  $\mathcal{B}$ , and for any element  $b \in \mathcal{B}$ , the principal filter  $\uparrow b$  is an ultrafilter if, and only if,  $b$  is an atom of  $\mathcal{B}$ . We refer to such ultrafilters as *principal*

*ultrafilters*, and to all of the other ultrafilters as *free ultrafilters*. In particular, if  $\mathcal{B}$  is finite, then every ultrafilter is principal, and therefore we recover  $\mathcal{B}$  entirely with the data of its ultrafilters, without topology.

- For any infinite set  $S$ ,  $\text{Cof}(S)$  is not an ultrafilter: for any subset  $T$  of  $S$  such that  $T$  and  $T^c$  are infinite, then neither  $T$  nor  $T^c$  belong to  $\text{Cof}(S)$ .

*Remark 1.9.* An important result, known as the *ultrafilter lemma* is that every proper filter is contained in an ultrafilter. This result requires Zorn's lemma in order to be proved. In particular, for every filter  $\mathcal{F}$  on a Boolean algebra  $\mathcal{B}$ ,

$$\mathcal{F} = \bigcap_{\substack{\alpha \text{ is an ultrafilter} \\ \mathcal{F} \subseteq \alpha}} \alpha.$$

A *filter basis*  $\mathfrak{B}$  is a non-empty collection of elements of  $\mathcal{B}$  which does not contain the bottom element and such that, for every  $b_1, b_2 \in \mathfrak{B}$  there exists  $b \in \mathfrak{B}$  such that  $b \leq b_1 \wedge b_2$ . A common way to describe an ultrafilter is to define a filter basis, and then to use the ultrafilter lemma in order to extend it into one of the (numerous) ultrafilters containing it.

**Definition 1.10** (Boolean spaces). A *Boolean space* is a topological space that is compact, Hausdorff, and that possesses a basis of clopen subsets. We denote by  $\text{Clop}(X)$  the Boolean algebra of clopen subsets of any Boolean space  $X$ , equipped with union, intersection and complement of subsets. We denote by **BStone** the category for which the objects are Boolean spaces and the arrows are continuous maps.

We are now ready to generalize Proposition 1.3 to arbitrary Boolean algebras.

**Theorem 1.11** (Stone duality for Boolean algebras, [67] Theorem 67). *There is a contravariant equivalence of categories between the category of Boolean algebras and the category of Boolean spaces.*

*Proof sketch.* We describe the two contravariant functors involved in the duality, the reader can check that they form a contravariant equivalence of categories. The contravariant functor  $\mathcal{S} : \mathbf{Bool} \rightarrow \mathbf{BStone}$  sends any Boolean algebra  $\mathcal{B}$  to the Boolean space  $\mathcal{S}(\mathcal{B})$ , and any Boolean algebra homomorphism  $h : \mathcal{B} \rightarrow \mathcal{B}'$  to the continuous map

$$\mathcal{S}(h) : \mathcal{S}(\mathcal{B}') \rightarrow \mathcal{S}(\mathcal{B}),$$

which sends every ultrafilter  $\gamma$  of  $\mathcal{B}'$  to

$$\mathcal{S}(h)(\gamma) := h^{-1}(\gamma) = \{b \in \mathcal{B} : h(b) \in \gamma\}.$$

The contravariant functor  $\text{Clop} : \mathbf{BStone} \rightarrow \mathbf{Bool}$  sends any Boolean space  $X$  to the Boolean algebra  $\text{Clop}(X)$ , and any continuous map  $f : X \rightarrow Y$  to the continuous map  $\text{Clop}(f) : \text{Clop}(Y) \rightarrow \text{Clop}(X)$ , which sends every clopen  $K$  of  $Y$  to  $f^{-1}(K)$ .  $\square$

In particular, for any Boolean algebra  $\mathcal{B}$ , the map  $(\hat{\cdot}) : \mathcal{B} \rightarrow \text{Clop}(\mathcal{S}(\mathcal{B}))$  which sends any element  $b$  in  $\mathcal{B}$  to the clopen subset  $\hat{b}$  is a Boolean isomorphism.

$$\begin{array}{ccc}
 \mathbf{Bool} & \xrightleftharpoons[\mathcal{S}(\cdot)]{Clop(\cdot)} & \mathbf{BStone} \\
 \uparrow & & \uparrow \text{pro-completion} \\
 \mathbf{Bool}_{fin} & \xrightleftharpoons[At(\cdot)]{\mathcal{P}(\cdot)} & \mathbf{Set}_{fin}
 \end{array}$$

Figure 8: Boolean Stone duality

We summarize these duality-theoretic relationships in Figure 8. At the bottom, we have the duality at the finite level, between finite Boolean algebras and finite sets. On the left is the fact that the category of finite Boolean algebras is a subcategory of the category of Boolean algebras. On the right, we use the fact that the category of Boolean spaces is equivalent to the pro-completion of the category of finite sets (see [38] Theorem VI, 2.3), and view the category of finite sets as a subcategory of the pro-completion of the category of finite sets. Finally, at the top, we have the duality between Boolean algebras and Boolean spaces.

*Remark 1.12.* In order to enforce this geometric perspective, we refer to the ultrafilters in the dual space as *the points* of the dual space. This terminology can also be perceived as analogous to the one used in linear algebra, where the points are linear maps. Denote by  $\mathbf{2}$  the two element Boolean algebra. For any Boolean algebra  $\mathcal{B}$ , there is a bijective correspondence between the set of Boolean algebra homomorphisms  $h : \mathcal{B} \rightarrow \mathbf{2}$  and the set of ultrafilters of  $\mathcal{B}$ . Indeed, on the one hand, consider the map which associates to each homomorphism  $h : \mathcal{B} \rightarrow \mathbf{2}$  the ultrafilter  $\gamma_h := h^{-1}(\{1\})$ . On the other hand, consider the inverse map, which associates to each ultrafilter  $\gamma$  of  $\mathcal{B}$  the homomorphism  $h_\gamma : \mathcal{B} \rightarrow \mathbf{2}$ , sending  $b \in \mathcal{B}$  to 1 if, and only if,  $b \in \gamma$ .

We end the subsection with the existence of coproduct for Boolean algebras and how they reflect as products on the topological side. We require this result, as the main Boolean space we are interested in in Chapter 2 corresponds to a subspace of the dual of a coproduct of Boolean algebras, see Proposition 2.3. For a proof of this result and more information on the categorical treatment of these structures, the reader can refer to [38], more specifically 2.12.

**Proposition 1.13.** *The category of Boolean algebras has coproducts. These are dual to product of Boolean spaces, which are calculated as in the category of topological spaces.*

### 1.1.2 The dual space of the powerset algebra

We focus on the duality theory in the particular case where the Boolean algebra we consider in Theorem 1.11 is the powerset algebra of a given set. Several additional observations can be made which will be useful in the next chapters.

The inclusion of the category of compact Hausdorff spaces into the category of topological spaces has a left adjoint, that we denote by  $\beta$ , which sends a general topological space to a compact Hausdorff topological space called its *Čech-Stone compactification*.

The Čech-Stone compactification of a space  $S$  can be defined by the following universal property: for any compact Hausdorff space  $X$  and any continuous map  $f : S \rightarrow X$ , there exists a unique continuous map  $g : \beta(S) \rightarrow X$  such that the following diagram commutes.

$$\begin{array}{ccc} S & \xrightarrow{\iota_S} & \beta(S) \\ & \searrow f & \downarrow g \\ & & X \end{array}$$

For any set  $S$ , the Stone dual of the powerset algebra  $\mathcal{P}(S)$  corresponds to the Čech-Stone compactification of the discrete space  $(S, \tau_{disc})$  (see [38] III, 2.1). Endowing  $S$  with the discrete topology,  $S$  can be embedded as a dense subspace of  $\beta(S)$  by considering the injective map  $\iota_S : S \rightarrow \beta(S)$  which sends any  $s$  in  $S$  to the principal ultrafilter  $\uparrow\{s\}$  (we use the abusive notation  $\uparrow s$ ).

Let us now consider two sets  $S$  and  $T$  and a map  $f : S \rightarrow T$ . In particular, we can consider the map  $\iota_T \circ f : S \rightarrow \beta(T)$ . Since  $\beta(T)$  is compact Hausdorff, by applying the universal property of the Čech-Stone compactification of  $S$ , there exists a unique continuous map

$$\beta f : \beta(S) \rightarrow \beta(T)$$

which extends  $f$ , and it is defined as follows: for any  $\alpha \in \beta(S)$ ,

$$\beta f(\alpha) = \{P \subseteq T : f^{-1}(P) \in \alpha\}.$$

We denote by  $\beta(S) \setminus S$  the closed subset of all free ultrafilters, that we also refer to as *the remainder* of  $\beta(S)$ . We often use the notation  ${}^*S := \beta(S) \setminus S$ .

**Lemma 1.14.** *For any set  $S$ , and any ultrafilter  $\alpha \in \beta(S)$ , we have that  $\alpha \in {}^*S$  if, and only if,  $\alpha$  contains all cofinite sets.*

*Proof.* Fix a set  $S$  and an ultrafilter  $\alpha \in \beta(S)$ . We prove the negation of this equivalence, that is  $\alpha$  is a principal ultrafilter if, and only if, there exists a cofinite set which does not belong to  $\alpha$ . For the left-to-right implication, suppose that  $\alpha$  is of the form  $\uparrow s$  for some  $s \in S$ . Then the set  $Q_s := S \setminus \{s\}$  is cofinite, and does not belong to  $\alpha$ . For the right-to-left implication, assume that there is a cofinite set  $Q$  of  $S$  which does not belong to  $\alpha$ . Then, since  $\alpha$  is an ultrafilter,  $Q^c$ , which is finite, does belong to  $\alpha$ . Since



an ultrafilter which contains a finite set is necessarily principal, we conclude that  $\alpha$  is principal.  $\square$

Let us say a few things about clopen subsets of  $\beta(S)$ . First, since  $(\cdot) : \mathcal{P}(S) \rightarrow \text{Clop}(\beta(S))$  is bijective, every clopen subset  $K$  of  $\beta(S)$  is of the form  $\widehat{Q}$ , for some  $Q \subseteq S$ . We can also prove ([78], Proposition 3.13) that every clopen subset of the remainder is of the form

$${}^*Q := \widehat{Q} \setminus Q := \{\alpha \in {}^*S : Q \in \alpha\},$$

for some infinite subset  $Q$  of  $S$ . We provide a few properties on clopen subsets of the remainder which will make computations in Chapter 2 easier.

**Lemma 1.15.** *For any set  $S$ , if  $Q_1$  and  $Q_2$  are two infinite subsets of  $S$ , then the following statements hold.*

- (1):  ${}^*Q_1 \subseteq {}^*Q_2$  if, and only if,  $Q_1 \setminus Q_2$  is finite.
- (2):  ${}^*Q_1 = {}^*Q_2$  if, and only if, the symmetric difference  $Q_1 \Delta Q_2$  is finite.
- (3):  ${}^*Q_1 \cap {}^*Q_2$  is non-empty if, and only if,  $Q_1 \cap Q_2$  is infinite.

*Proof.* See [78], Proposition 3.14.  $\square$

We conclude this section by proving a few topological properties that will be used regularly in Chapter 2. First, we prove a classical lemma about finite partitions of  $S$ .

**Lemma 1.16.** *For any ultrafilter  $\alpha \in \beta(S)$ , and any finite partition  $(Q_1, \dots, Q_n)$  of  $S$  for some  $n \geq 1$ , there exists a unique  $k \in \{1, \dots, n\}$  such that  $Q_k \in \alpha$ .*

*Proof.* Fix  $\alpha \in \beta(S)$ , and a finite partition  $(Q_1, \dots, Q_n)$  of  $S$ , for some  $n \geq 1$ . For the existence, we proceed by finite induction. In the case  $n = 1$ , the result is obvious. If  $(Q, Q^c)$  is a partition of  $S$ , then since  $\alpha$  is an ultrafilter, there is exactly one element in  $\{Q, Q^c\}$  which is in  $\alpha$ . Now, fix  $n \geq 3$  and a finite partition  $(Q_1, \dots, Q_n)$  of  $S$ . Assume that we proved the statement for a finite partition with  $n - 1$  elements. If  $Q_1$  is in  $\alpha$ , then we are done. Otherwise,  $Q_1^c$  is in  $\alpha$ , but since we have a finite partition of  $S$ ,  $Q_1^c = \bigcup_{i=2}^n Q_i$ . By applying the induction hypothesis, there exists  $k \in \{2, \dots, n\}$  such that  $Q_k \in \alpha$ , which allows us to conclude. For the unicity, if there existed  $Q_k$  and  $Q_{k'}$ , with  $k \neq k'$ , both in  $\alpha$ , then their intersection would also be in  $\alpha$ . Since we have a finite partition of  $S$ , this intersection is empty, and therefore  $\alpha$  cannot be an ultrafilter.  $\square$

*Remark 1.17.* Note that the argument used in this proof works for any ultrafilter  $x$  in a Boolean algebra  $\mathcal{B}$ , the assumption made here that  $\mathcal{B} = \mathcal{P}(S)$  is not needed. However, we only use this case later on.

Finally, we prove a technical lemma that holds for any Boolean space, and appears as a generalization of the fact that these spaces are totally disconnected. Its relevance will become more clear in Chapter 2, Proposition 2.33.

**Lemma 1.18.** *Let  $X$  be a Boolean space. For any finite family of distinct points  $\alpha_1, \dots, \alpha_\ell \in X$ , there exists a family of pairwise disjoint clopen subsets  $K_1, \dots, K_\ell$  of  $X$  such that, for every  $i \in \{1, \dots, \ell\}$ ,  $\alpha_i \in K_i$ .*

*Proof.* We prove the statement by induction on  $\ell \geq 1$ . In the case  $\ell = 1$ , the statement is obvious. In the case  $\ell = 2$ , the statement is true since  $X$  is a Boolean space, that is, totally disconnected, and compact Hausdorff. Now, assume that the statement holds for a fixed  $\ell \geq 3$ . Fix  $\alpha_1, \dots, \alpha_{\ell+1} \in X$  a family of  $\ell + 1$  distinct points. By the induction hypothesis, there exists a family of  $\ell$  pairwise disjoint clopen subsets  $K_1, \dots, K_\ell$  such that, for every  $i \in \{1, \dots, \ell\}$ ,  $\alpha_i \in K_i$ . Now, as we mentioned for the binary case, any two distinct points in a Boolean space can be separated by two clopen subsets. Therefore, for every  $i \in \{1, \dots, \ell\}$ , we have two clopen subsets  $L_i, L_i^{\ell+1}$  which contain, respectively,  $\alpha_i$  and  $\alpha_{\ell+1}$ . Now, setting

$$R_{\ell+1} := \bigcap_{i=1}^{\ell} L_i^{\ell+1}$$

and, for every  $i \in \{1, \dots, \ell\}$ ,

$$R_i := K_i \cap L_i,$$

the family of pairwise disjoint clopen subsets  $(R_1, \dots, R_{\ell+1})$  is, by construction, such that for every  $i \in \{1, \dots, \ell + 1\}$ ,  $\alpha_i \in R_i$ .  $\square$

**Proposition 1.19.** *Let  $S$  be a set. For any finite family of distinct points  $\alpha_1, \dots, \alpha_\ell \in \beta(S)$ , there exists a family of pairwise disjoint subsets  $Q_1, \dots, Q_\ell$  of  $S$  such that, for every  $i \in \{1, \dots, \ell\}$ ,  $Q_i \in \alpha_i$ .*

*In particular, if  $C$  is an infinite subset of  $\beta(S)$ , then, for every  $\ell \in \mathbb{N}$ , there exists a family  $Q_1, \dots, Q_\ell$  of pairwise disjoint subsets of  $S$  such that, for every  $i \in \{1, \dots, \ell\}$ ,  $C \cap \widehat{Q_i}$  is non-empty.*

*Proof.* This is a direct consequence of Lemma 1.18.  $\square$

## 1.2 Modal algebra and the Vietoris functor

In [76], Vietoris introduced a generalization of the notion of Hausdorff metrics on any compact Hausdorff space: the so-called Vietoris hyperspace of a topological space.

**Definition 1.20** (Vietoris hyperspace of a Boolean space [76]). For any Boolean space  $X$ , we denote by  $\mathcal{V}(X)$  the set of closed subsets of  $X$ . We endow it with the topology generated by the sets of the form

$$\square K := \{C \in \mathcal{V}(X) : C \subseteq K\} \text{ and } \diamond K := \{C \in \mathcal{V}(X) : C \cap K \neq \emptyset\},$$

for every clopen subset  $K$  of  $X$ , and we refer to this topological space as the *Vietoris hyperspace* of  $X$ .

We make a few remarks on this construction. First, observe that for any clopen subset  $K$  of  $X$ ,

$$\Diamond K = (\Box K^c)^c,$$

therefore elements of the form  $\Diamond K$  can be replaced by elements of the form  $(\Box K')^c$ , where  $K' = K^c$ . Therefore, we may also define  $\mathcal{V}(X)$  by taking  $\{\Box K, (\Box K)^c : K \in \text{Clop}(X)\}$  as a basis. We also note that  $\Box$  is meet-preserving, while  $\Diamond$  is join preserving. In particular, the family of the clopen subsets of the form

$$\langle K, K_1, \dots, K_n \rangle := \Box K \cap \bigcap_{i=1}^n \Diamond K_i,$$

where  $n \geq 1$  and  $K, K_1, \dots, K_n$  are clopen subsets of  $X$  provides a basis for the Vietoris topology.

*Remark 1.21.* Fix a finite sequence of clopen subsets  $K, K_1, \dots, K_n \subseteq X$ . A simple, yet important, observation is that, for every clopen  $K' \subseteq X$  such that  $K \subseteq K'$ , we have

$$\langle K, K_1, \dots, K_n \rangle \subseteq \langle K', K_1, \dots, K_n \rangle.$$

The same way, for every finite sequence of clopen subsets  $K'_1, \dots, K'_n \subseteq X$  such that, for every  $i \in \{1, \dots, n\}$ ,  $K_i \subseteq K'_i$ , we have

$$\langle K, K_1, \dots, K_n \rangle \subseteq \langle K, K'_1, \dots, K'_n \rangle.$$

The Vietoris hyperspace  $\mathcal{V}(X)$  is defined for any topological space  $X$ , replacing the compact opens of your Definition 2.4.1 with a basis of the topology of  $X$ . However, when  $X$  is a Boolean space, the Vietoris hyperspace of  $X$  is also a Boolean space, see [47], Theorem 4.9. Note that, since it is totally disconnected, it is  $T_1$ , and thus singletons are closed (this remark will be of use in Proposition 2.33). Also, note that any clopen of  $\mathcal{V}(X)$  is a compact space and thus can be written as a finite union of clopens of the form  $\langle K, K_1, \dots, K_n \rangle$ , where  $n \in \mathbb{N}$  and  $K, K_1, \dots, K_n$  are clopen subsets of  $X$ .

*Remark 1.22.* For any Boolean algebra  $\mathcal{B}$ , any set  $S$  and any map  $f : S \rightarrow \mathcal{V}(\mathcal{S}(\mathcal{B}))$  the universal property of Cech-Stone compactification states that  $f$  admits a unique continuous extension  $g : \beta(S) \rightarrow \mathcal{V}(\mathcal{S}(\mathcal{B}))$ , defined by sending any ultrafilter  $\alpha \in \beta(S)$  to

$$g(\alpha) := \bigcap_{\substack{b \in \mathcal{B} \\ \{s \in S : f(s) \subseteq \widehat{b}\} \in \alpha}} \widehat{b}.$$

This remark will be useful in order to prove Proposition 2.7

Considering  $\mathcal{B}$  the Boolean algebra dual to  $X$ , we can understand the Vietoris hyperspace on  $X$  by equipping the set of filters of  $\mathcal{B}$  with an appropriate topology. Since our approach relies on a more topological understanding of problems, we chose to conduct most of our reasoning in terms of closed subsets. However, it should be noted

that it is only a matter of preference, and that one could formulate all of the results from Chapter 2 involving closed subsets in terms of filters instead.

**Proposition 1.23.** *Let  $\mathcal{B}$  be a Boolean algebra and  $X = \mathcal{S}(\mathcal{B})$  its dual space. The Vietoris hyperspace of  $X$  is homeomorphic to the space of filters of  $\mathcal{B}$ ,*

$$\mathcal{V}(X) \simeq \text{Filt}(\mathcal{B}),$$

where  $\text{Filt}(\mathcal{B})$  is endowed with the topology generated by the clopen sets of the form  $[b]$  and  $[b]^c$ , where for every  $b \in \mathcal{B}$ ,

$$[b] := \{\mathcal{F} \in \text{Filt}(\mathcal{B}) : b \in \mathcal{F}\}.$$

*Proof.* For any Boolean algebra, we give a one-to-one correspondence between the set of its filters and the closed subsets of its dual space, and we prove that this correspondence is a homeomorphism. Since we will have a bijection between compact Hausdorff spaces, to prove that we have a homeomorphism it is enough to prove that one of them is continuous. Consider the map  $\varphi : \text{Filt}(\mathcal{B}) \rightarrow \mathcal{V}(\mathcal{S}(\mathcal{B}))$  which sends a filter  $\mathcal{F}$  to the intersection of the clopens corresponding to the elements of the filter,

$$\varphi(\mathcal{F}) := \bigcap_{b \in \mathcal{F}} \widehat{b},$$

and the map  $\psi : \mathcal{V}(\mathcal{S}(\mathcal{B})) \rightarrow \text{Filt}(\mathcal{B})$  which sends a closed subset  $C$  of the dual space to the filter of all elements of the Boolean algebra for which the corresponding clopen contains  $C$ ,

$$\psi(C) := \{b \in \mathcal{B} : C \subseteq \widehat{b}\}.$$

We prove that these two maps are mutually inverse functions. First, by definition of  $\varphi$  and  $\psi$ , we obtain for any  $\mathcal{F} \in \text{Filt}(\mathcal{B})$  and any  $C \in \mathcal{V}(\mathcal{S}(\mathcal{B}))$  that  $\mathcal{F} \subseteq \psi(\varphi(\mathcal{F}))$  and that  $C \subseteq \varphi(\psi(C))$ .

We prove that  $\varphi(\psi(C)) \subseteq C$ . In order to do so, we prove the contrapositive. Let  $x$  be a point that does not belong to  $C$ . Since  $C$  is a closed subset of  $\beta(S)$ ,  $C^c$  is an open subset of  $\beta(S)$ . As we know that the family of clopen of the form  $\widehat{b}$ , where  $b$  ranges over  $\mathcal{B}$ , forms a basis of  $\beta(S)$ , we can pick  $b \in \mathcal{B}$  such that  $x \in \widehat{b}$  and  $\widehat{b} \cap C$  is empty. In particular,  $\neg b$  is such that  $C \subseteq \widehat{\neg b}$ , but  $x \notin \widehat{\neg b}$ , so that  $x \notin \varphi(\psi(C))$ , allowing us to conclude.

We prove that  $\psi(\varphi(\mathcal{F})) \subseteq \mathcal{F}$ . Let  $a$  be any element of the Boolean algebra such that  $\varphi(\mathcal{F}) \subseteq \widehat{a}$ . Then,  $\varphi(\mathcal{F})$  is disjoint from  $\widehat{\neg a}$ , but  $\varphi(\mathcal{F})$  is an intersection of closed sets. By the formulation of compactness in terms of intersections, this means that there exist  $b_1, \dots, b_n \in \mathcal{F}$  such that  $\bigcap_{i=1}^n \widehat{b_i} \subseteq \widehat{a}$ . Now, since  $\widehat{\cdot} : \mathcal{B} \rightarrow \text{Clop}(\mathcal{S}(\mathcal{B}))$  is an injective homomorphism, we have that  $\bigcap_{i=1}^n b_i \subseteq a$ , and we conclude that  $a \in \mathcal{F}$ .

Now, since inverse image preserves Boolean operations, so in order to prove that  $\psi$

is continuous, it is enough to check that  $\psi^{-1}([b])$  is a clopen subset of  $\mathcal{V}(\mathcal{S}(\mathcal{B}))$ , for any  $b$  in  $\mathcal{B}$ . Since

$$\begin{aligned}\psi^{-1}([b]) &= \{C \in \mathcal{V}(\mathcal{S}(\mathcal{B})) : \psi(C) \in [b]\} \\ &= \{C \in \mathcal{V}(\mathcal{S}(\mathcal{B})) : C \subseteq \widehat{b}\} \\ &= \square \widehat{b},\end{aligned}$$

where  $\widehat{b}$  is indeed clopen, we are able to conclude.  $\square$

For any set  $S$ , by applying Proposition 1.23 to the Boolean algebra  $\mathcal{P}(S)$ , we obtain the following result.

**Corollary 1.24.** *For any set  $S$ , the Vietoris hyperspace of the Čech-Stone compactification of  $S$  is homeomorphic to the space of filters on  $S$ ,*

$$\mathcal{V}(\beta(S)) \simeq \text{Filt}(S).$$

For any Boolean algebra  $\mathcal{B}$ , we denote by  $C_{\mathcal{F}}$  the closed subset of  $\mathcal{S}(\mathcal{B})$  corresponding to a filter  $\mathcal{F} \in \text{Filt}(\mathcal{B})$  under this correspondence, and reciprocally, we denote by  $\mathcal{F}_C$  the filter corresponding to a closed subset  $C$ . In the particular case where  $\mathcal{B} = \mathcal{P}(S)$  for some set  $S$ , we have that, for any  $C \in \mathcal{V}(X)$ ,

$$\mathcal{F}_C = \{Q \subseteq S : \forall \alpha \in C, Q \in \alpha\} = \bigcap_{\alpha \in C} \alpha$$

and for any  $\mathcal{F} \in \text{Filt}(S)$ ,

$$C_{\mathcal{F}} = \{\alpha \in \beta(S) : \mathcal{F} \subseteq \alpha\}.$$

**Example 1.25.** Fix a set  $S$ . We provide different instances of the correspondence introduced in Corollary 1.24.

- Fix a subset  $Q$  of  $S$ . The filter of  $S$  corresponding to  $\widehat{Q}$ , is the principal filter containing  $Q$ , since

$$\mathcal{F}_{\widehat{Q}} = \{P \subseteq S : \widehat{Q} \subseteq \widehat{P}\} = \{P \subseteq S : Q \subseteq P\} = \uparrow Q.$$

The closed subset of  $\beta(S)$  corresponding to  $\uparrow Q$  is  $\widehat{Q}$ , since

$$C_{\uparrow Q} = \bigcap_{P \in \uparrow Q} \widehat{P} = \bigcap_{Q \subseteq P} \widehat{P} = \widehat{Q}.$$

- The closed subset of  $\beta(S)$  corresponding to  $\text{Cof}(S)$ , the filter of all of the cofinite subsets of  $S$  is the remainder of  $S$ , and vice-versa. Indeed, by Proposition 1.23, we have

$$C_{\text{Cof}(S)} := \{\alpha \in \beta(S) : \forall Q \in \text{Cof}(S), Q \in \alpha\}$$

which is equal to  ${}^*S$  by Lemma 1.14, and

$$\mathcal{F}_{*S} = \{Q \subseteq S : {}^*S \subseteq \widehat{Q}\} = \{Q \subseteq S : {}^*S \subseteq {}^*Q\}$$

which is equal to  $\text{Cof}(S)$  by Lemma 1.15 (1).

More generally, for any subset  $Q$  of  $S$ , the closed subset of  $\beta(S)$  corresponding to the filter

$$\uparrow Q \cap \text{Cof}(S) := \{P \subseteq S : Q \subseteq P \text{ and } P^c \text{ is finite}\}$$

is  ${}^*Q$  and vice-versa. Indeed, by Proposition 1.23, we have

$$C_{\uparrow Q \cap \text{Cof}(S)} := \bigcap_{Q' \in \uparrow Q \cap \text{Cof}(S)} \widehat{Q'} = \bigcap_{Q' \in \uparrow \text{Cof}(S)} \widehat{Q'} \cap \widehat{Q}$$

which is equal to  ${}^*S \cap Q$ , that is,  ${}^*Q$ , and

$$\mathcal{F}_{\uparrow Q \cap \text{Cof}(S)} = \{Q' \subseteq S : {}^*Q \subseteq \widehat{Q'}\}$$

which is equal to  $\uparrow Q \cap \text{Cof}(S)$  by Lemma 1.15 (1).

We now define a notion of product for two filters of  $\mathcal{P}(S)$ , which will be a major example needed in Section 2.4.

**Definition 1.26.** Fix a set  $S$ . We define the  $\mathcal{P}(S^2)$ -product of two filters  $\mathcal{F}_1, \mathcal{F}_2 \in \text{Filt}(S)$  as the filter

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \uparrow \{P_1 \times P_2 : P_1 \in \mathcal{F}_1, P_2 \in \mathcal{F}_2\} \in \text{Filt}(S^2).$$

This filter corresponds, by Proposition 1.23, to the closed subset of  $\beta(S^2)$

$$C_{\mathcal{F}_1 \otimes \mathcal{F}_2} := \bigcap_{\substack{P_1 \in \mathcal{F}_1 \\ P_2 \in \mathcal{F}_2}} \widehat{P_1 \times P_2},$$

which motivates the following definition.

**Definition 1.27.** Fix a set  $S$ . We define the  $\beta(S^2)$ -product of two closed subsets  $C_1, C_2 \in \mathcal{V}(\beta(S))$  as

$$C_1 \otimes C_2 := \bigcap_{\substack{C_1 \subseteq \widehat{P_1} \\ C_2 \subseteq \widehat{P_2}}} \widehat{P_1 \times P_2}.$$

This terminology was implemented in order to avoid confusion with the product topology. However, since this is the only kind of product that we use in this thesis, we choose to abbreviate it as product. Note that the projections maps  $\pi_1, \pi_2 : S^2 \rightarrow S$  can be lifted into two continuous maps  $\beta\pi_1, \beta\pi_2 : \beta(S^2) \rightarrow \beta(S)$ , yet these maps are not bijective and thus it is not possible, in general, to retrieve an ultrafilter  $\alpha \in \beta(S^2)$  by only looking at its two projections.

**Example 1.28.** Fix a set  $S$ . We provide different instances of product of closed subsets of  $\beta(S)$ .

- In the case of two clopen subsets,  $\widehat{Q_1}, \widehat{Q_2} \subseteq \beta(S)$ , we have

$$\widehat{Q_1} \otimes \widehat{Q_2} = \bigcap_{\substack{Q_1 \subseteq P_1 \\ Q_2 \subseteq P_2}} \widehat{P_1 \times P_2} = \widehat{Q_1 \times Q_2}.$$

- In the case where the closed subset is the remainder, we have

$$^*S \otimes ^*S = \bigcap_{\substack{F_1, F_2 \subseteq S \\ F_1, F_2 \text{ finite}}} (S \setminus F_1) \times (S \setminus F_2).$$

The Vietoris construction can be seen as a functor  $\mathcal{V} : \mathbf{BStone} \rightarrow \mathbf{BStone}$  on the category of Boolean spaces and continuous functions. Indeed, if  $f : X \rightarrow Y$  is continuous, then so is the continuous map  $\mathcal{V}(f) : \mathcal{V}(X) \rightarrow \mathcal{V}(Y)$  which sends a closed subset  $C$  of  $X$  to  $f(C)$ . We would like to complete the following commutative diagram.

$$\begin{array}{ccc} & S & \\ \text{?} \curvearrowright & \xrightarrow{\quad} & \text{Stone} \curvearrowright \nu \\ & \xleftarrow{\quad \text{Clop} \quad} & \end{array} \quad BA$$

In order to do so, we need to define a functor  $M : \mathbf{Bool} \rightarrow \mathbf{Bool}$  which could be seen as the dual of  $\mathcal{V} : \mathbf{BStone} \rightarrow \mathbf{BStone}$ . This functor sends any Boolean algebra to what is called its corresponding formal *modal algebra*. We could summarize modal algebra by saying that, just as Boolean algebras are models of classical logic, modal algebras provide models of propositional modal logic. The reader intested in a complete introduction to the framework of modal logic and its uses can refer to [8], and to [41]. More specific results about the relationships between Vietoris topology, modal logic and coalgebras are also available. This topic is especially relevant to this thesis, since universal quantification acts like a box operation in our setting.

**Definition 1.29** (Modal algebra). A *modal algebra* is a pair  $(\mathcal{B}, \Box)$  where  $\mathcal{B}$  is a Boolean algebra and  $\Box : \mathcal{B} \rightarrow \mathcal{B}$  satisfies the following properties:  $\Box 1 = 1$  and for every  $b_1, b_2$  in  $\mathcal{B}$ ,  $\Box(b_1 \wedge b_2) = \Box b_1 \wedge \Box b_2$ .

For any Boolean algebra  $\mathcal{B}$ , we denote by  $M\mathcal{B}$  the free Boolean algebra over the set of formal generators  $\{\Box b : b \in \mathcal{B}\}$ , with the following relations:  $\Box 1 = 1$  and for every  $b_1, b_2$  in  $\mathcal{B}$ ,  $\Box(b_1 \wedge b_2) = \Box b_1 \wedge \Box b_2$ . This is the free Boolean algebra over the semilattice reduct of  $\mathcal{B}$ . What this means concretely is that  $M\mathcal{B}$  can be characterized as the Boolean algebra expansion of  $\mathcal{B}$  with the property that, for any meet-preserving function between Boolean algebras  $h : \mathcal{B} \rightarrow \mathcal{B}'$ , there is a unique Boolean algebra homomorphism  $\bar{h} : M\mathcal{B} \rightarrow M\mathcal{B}'$  which extends  $h$ . We could have defined  $M\mathcal{B}$  in a similar

fashion by introducing the generators  $\Diamond b$ , for any  $b \in \mathcal{B}$ , with the relations  $\Diamond 0 = 0$  and for any  $b_1, b_2$  in  $\mathcal{B}$ ,  $\Diamond(b_1 \vee b_2) = \Diamond b_1 \vee \Diamond b_2$ . For any  $b \in \mathcal{B}$ , the relation  $\Box b = \neg(\Diamond \neg b)$  holds. It is already fairly transparent that this mirrors the topological structure provided by the Vietoris hyperspace, on the algebraic level.

One can use the Vietoris construction in order to understand the dual space of the modal algebra built on a Boolean algebra  $\mathcal{B}$ .

**Proposition 1.30.** *For any Boolean algebra  $\mathcal{B}$ , the dual space of  $M\mathcal{B}$  is homeomorphic to the Vietoris hyperspace of the dual space of  $\mathcal{B}$ .*

*Proof.* See [41], Fact 1. □

An important notion we will require in Chapter 2 is the notion of content of a closed subset of  $\beta(S)$ . Basically, it consists in only looking at the points in the closed subset which correspond to principal ultrafilters.

**Definition 1.31.** For any set  $S$ , the content of a closed subset  $C$  of  $\beta(S)$  is

$$\text{Cont}(C) := C \cap S.$$

Note that the content of a closed subset of  $\beta(S)$  may very well be empty in general.

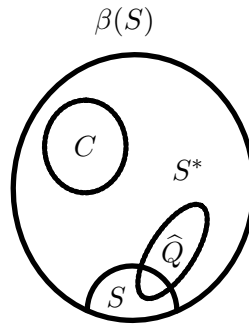


Figure 9: The space of ultrafilters  $\beta(S)$ . Clopen subsets must contain elements of  $S$ , whilst closed subsets do not need to.

For the reader who would rather prefer to reason in terms of filters, the corresponding definition is the following. The content of a filter  $\mathcal{F} \in \text{Filt}(S)$ , can be defined as the set of elements of  $S$  appearing in every set of the filter: that is,

$$\text{Cont}(\mathcal{F}) := \bigcap \mathcal{F} = \bigcap_{Q \in \mathcal{F}} Q.$$



**Example 1.32.** We compute the content of the closed subsets introduced in Example 1.25. Fix a set  $S$ .

- For any  $Q \subseteq S$ , the content of the closed subset  $\widehat{Q}$  of  $\beta(S)$  is  $\widehat{Q} \cap S = Q \cap S = Q$ .
- The content of the remainder  ${}^*S$  is the empty set. This is an example of a closed subset of  $\beta(S)$  with an empty content, but which is not the empty set. More generally, for any  $Q \subseteq S$ , the content of the closed subset  ${}^*Q$  of  $\beta(S)$  is the empty set.
- For any closed subsets  $C_1, C_2$  of  $\beta(S)$ , we have that

$$\begin{aligned}
 \text{Cont}(C_1 \otimes C_2) &= \left( \bigcap_{\substack{C_1 \subseteq \widehat{P_1} \\ C_2 \subseteq \widehat{P_2}}} \widehat{P_1 \times P_2} \right) \cap S^2 \\
 &= \bigcap_{\substack{C_1 \subseteq \widehat{P_1} \\ C_2 \subseteq \widehat{P_2}}} (\widehat{P_1 \times P_2} \cap S^2) \\
 &= \bigcap_{\substack{C_1 \subseteq \widehat{P_1} \\ C_2 \subseteq \widehat{P_2}}} \{(s_1, s_2) \in S^2 : s_1 \in P_1 \text{ and } s_2 \in P_2\} \\
 &= (C_1 \cap S) \times (C_2 \cap S) \\
 &= \text{Cont}(C_1) \times \text{Cont}(C_2).
 \end{aligned}$$

### 1.3 Formal languages and recognition

In this section, we introduce the basics of formal language theory, and the notion of *language recognition* by finite automata, and its algebraic counterpart via finite monoids. We then explain how it is possible to link it to duality theory from Section 1.1 and finally provide a few examples of this connection. For a more thorough introduction to automata theory, we recommend [42] and [63], and for the algebraic treatment of language recognition we refer to [71] and [54].

Throughout the rest of the thesis, we fix a finite alphabet  $A$ , and we refer to the elements of  $A^*$ , the free monoid over  $A$ , as the *finite words* on  $A$ . We denote by  $|w|$  the *length* of a finite word  $w = w_0 \dots w_{|w|-1} \in A^*$ , where for every  $i \in \{0, \dots, |w| - 1\}$ ,  $w_i$  is in  $A$ . Finally, we denote by  $|w|_a$  the number of occurrences of the letter  $a$  in the word  $w$ . A *language*  $L$  is a subset of  $A^*$ , and since  $\mathcal{P}(A^*)$  is a Boolean algebra for union, intersection and complement, these operations are naturally defined on languages. One last operation on languages that is extremely useful is the quotient by a word.

**Definition 1.33.** For any language  $L \subseteq A^*$  and any word  $v \in A^*$ , we define the *left quotient*  $v^{-1}.L$  and the *right quotient*  $L.v^{-1}$  as the languages

$$v^{-1}.L := \{w \in A^* : v.w \in L\}$$

and

$$L.v^{-1} := \{w \in A^* : w.v \in L\}.$$

We now introduce the notion of recognition by finite automaton, which plays a major role in the comprehension of finite languages.

**Definition 1.34.** A *finite automaton* is a tuple  $\mathcal{A} = (Q, \delta, I, F)$ , where

- $Q$  is a finite set that we refer to as the *set of states* of the automaton;
- $\delta \subseteq Q \times A \times Q$  is a relation that we refer to as the *transition relation* of the automaton;
- $I, F \subseteq Q$  are referred to as the set of *initial states*, and the set of *final states* of the automaton, respectively.

A word  $w \in A^*$  with length  $n \in \mathbb{N}_{>0}$  is accepted by an automaton  $\mathcal{A} = (Q, \delta, I, F)$  if there exists  $q_0, \dots, q_n \in Q$ , where  $q_0 \in I$ ,  $q_n \in F$  and for every  $i \in \{0, \dots, n-1\}$ ,  $(q_i, w_i, q_{i+1}) \in \delta$ . We say that a language is recognized by an automaton if the automaton accepts every word in the language. A language  $L \subseteq A^*$  is called *regular* if there exists a finite automaton which recognizes it. We denote by  $\text{Reg}(A^*)$  the subset of  $A^*$  of all regular languages. The set of all regular languages is a Boolean subalgebra of  $\mathcal{P}(A^*)$ . One can prove that it is closed under left and right quotients and, that the set  $\{u^{-1}.L.v^{-1} : u, v \in A^*\}$  is finite (see [63] Chapter 1, Section 2).

**Example 1.35.** If the alphabet we consider is  $A = \{a, b\}$ , then the language  $A^*aA^*$  is regular, since it is recognised, for instance, by the automaton where  $Q = \{q_0, q_1\}$ ,  $\delta = \{(q_0, b, q_0), (q_0, a, q_1), (q_1, a, q_1), (q_1, b, q_1)\}$ ,  $I = \{q_0\}$  and  $F = \{q_1\}$ .

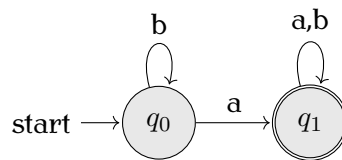


Figure 10: An automaton that recognises the language  $A^*aA^*$ , where  $A = \{a, b\}$ .

The algebraic approach to language theory consists in studying a notion of recognition based not on finite automaton, but on finite monoids. This is a very helpful point of view when studying languages since one can use the many ideas and results available in monoid theory.

**Definition 1.36.** A language  $L \subseteq A^*$  is said to be recognised by a monoid homomorphism  $h : A^* \rightarrow M$  if there exists a subset  $P \subseteq M$  such that

$$h^{-1}(P) = L.$$

More generally, a Boolean algebra of languages is said to be recognized by  $h$  if  $h$  recognizes every language in  $\mathcal{B}$ .

This notion is equivalent to the notion of recognition by finite automaton that we introduced initially.

**Proposition 1.37.** *For every language  $L \subseteq A^*$ , the following conditions are equivalent.*

(1):  *$L$  is recognized by a finite automaton.*

(2):  *$L$  is recognized by a monoid homomorphism into a finite monoid.*

*Proof sketch.* For (1) implies (2), consider a language  $L$  recognized by a finite automaton  $\mathcal{A} = (Q, \delta, I, F)$ . We define  $M_{\mathcal{A}}$  as the monoid of binary relations  $R \subseteq Q^2$  on  $Q$ , where the monoid law is the composition of relations and the identity is the identity relation. One can prove that the map  $h : A^* \rightarrow M_{\mathcal{A}}$  which sends any finite word  $w \in A^*$  to

$$\{(q, q') : \text{there exists a path in } \mathcal{A} \text{ labelled } w \text{ from } q \text{ to } q'\}$$

is a monoid morphism, and that, setting  $P := \{R \subseteq Q^2 : R \cap (I \times F) \neq \emptyset\}$  we have that  $h^{-1}(P) = L$ , which proves that  $h$  recognizes  $L$ .

For (2) implies (1), consider a language  $L$  such that there exists a monoid morphism  $h : A^* \rightarrow M$ , where  $M$  is a finite monoid, which recognizes  $L$ . We define an automaton  $\mathcal{A}_M = (Q_M, \delta_M, I_M, F_M)$ , that recognizes  $L$ , as follows. We set  $Q_M := M, I_M := \{1_M\}, F_M := h(M)$  and  $\delta_M := \{(m, a, m.h(a)) : m \in M, a \in A\}$ . One can now prove that this automaton recognizes the language  $L$ .

□

There are infinitely many finite monoids that are able to recognize a given language  $L \in \text{Reg}(A^*)$ . However, there exists a monoid  $M_L$  which is “minimal” in the following sense:  $M_L$  recognizes  $L$ , and for every finite monoid  $M$  that recognizes  $L$ ,  $M_L$  is a quotient of a submonoid of  $M$ . We refer to this minimal recognizer as the *syntactic monoid*, and to  $h$  as the *syntactic homomorphism*. In particular, a language is recognizable if, and only if, its syntactic monoid is finite. Explicitly, the syntactic monoid of a language  $L$  is the quotient of  $A^*$  under the *syntactic congruence*  $\sim_L$  which is defined as follows. For any  $w, w' \in A^*$ ,  $w \sim_L w'$  if, and only if

$$\forall u, v \in A^*, (w \in u^{-1}.L.v^{-1} \iff w' \in u^{-1}.L.v^{-1}).$$

It is now time for us to explain how the framework of duality theory we introduced in Section 1.1 can be related to the theory of language recognition. Fix a language  $L \in \text{Reg}(A^*)$ , and consider the Boolean algebra

$$\mathcal{B}_L := \langle \{u^{-1}.L.v^{-1} : u, v \in A^*\} \rangle_{BA}.$$

As we mentioned previously, the generating set for this Boolean algebra is finite and thus so is  $\mathcal{B}_L$ . By Proposition 1.3, the embedding  $\mathcal{B}_L \hookrightarrow \mathcal{P}(A^*)$  is dual to a surjection  $A^* \twoheadrightarrow \text{At}(\mathcal{B}_L)$ , that is, an equivalence relation on  $A^*$ , such that the equivalence class of

$w \in A^*$  is

$$\bigcap_{u.w.v \in L} u^{-1}.L.v^{-1} \cap \bigcap_{u.w.v \notin L} (u^{-1}.L.v^{-1})^c.$$

These equivalence classes are exactly the equivalence classes of  $\sim_L$ , the syntactic congruence of  $L$ , and thus, correspond to the elements of the syntactic monoid  $M_L$  of  $L$ .

This short argument already justifies the pertinence of looking at recognition from a duality theoretic point of view: discrete duality allowed us to retrieve the notion of recognition by finite monoids. Even more interesting: if we enrich the structure on  $\text{Reg}(A^*)$  with a structure of residuation subalgebra of  $\mathcal{P}(A^*)$ , by considering the operations left and right quotient by languages, and studied the duality theoretic consequences of this change, we could directly generalise this argument to Boolean algebras of languages, closed under quotients, that are not necessary regular any more, cf [31].

*Remark 1.38.* This last observation is what led to the introduction of the formalism of *Boolean algebras with an internal monoids*, or *BiMs*. When it comes to studying Boolean algebras closed under quotients, BiMs constitute a more pertinent class of mathematical objects than finite monoids, since they also function in the non-regular setting, cf [30] and [31]. Many different directions can be followed in order to extend the tools available for regular languages to more general settings. In [9], the monoid morphisms are replaced by  $T$ -algebras, where  $T$  is a monad on a certain category. The framework of monads can be used to described algebraic approaches to other data structures than finite words, such as trees for instance. Another example is [17], where languages and their acceptors are replaced by functors between input categories (specifying the type of the languages and of the acceptors) and output categories (specifying the type of outputs).

In the more general case of an arbitrary Boolean algebra of languages however, no supplementary structure on the dual space is available, and therefore the topological analogue for the minimal automaton is simply the dual space. This is the case that will be at the center of our study in this thesis, which justifies us only paying attention to the topological structure of the dual space in Chapter 2.

**Definition 1.39.** A language  $L \subseteq A^*$  is recognized by the continuous map  $f : \beta(A^*) \rightarrow X$ , where  $X$  is a Boolean space, if there exists a clopen  $K \subseteq X$  such that

$$\hat{L} = f^{-1}(K).$$

More generally, a Boolean algebra of languages  $\mathcal{B}$  is recognized by  $f$  if, for every  $L \in \mathcal{B}$ ,  $L$  is recognized by  $f$ .

We refer to the dual space as the *minimal recognizer* of the Boolean algebra of languages we consider. In the case where the language we consider is regular,  $X = M$

is a finite monoid and the map  $f$  is a monoid homomorphism, Definition 1.39 coincides with the notion of language recognition by finite monoids.

**Example 1.40.** We give a few examples of computation of dual spaces for some Boolean algebras of languages.

- The dual space of  $\text{Reg}(A^*)$  is the topological space that we denote by  $\widehat{A^*}$  and refer to as the set of *profinite words*, obtained as the completion of  $A^*$  for the profinite metric  $d$ . We give an overview of this space, for more details on this structure the reader can refer to [?] and [1]. First, we say that a finite monoid  $M$  separates two finite words  $u, v \in A^*$  if there exists a monoid homomorphism  $\varphi : A^* \rightarrow M$  such that  $\varphi(u) \neq \varphi(v)$ . We then set, for any  $u, v \in A^*$ ,

$$r(u, v) := \min\{|M| : M \text{ is a finite monoid that separates } u \text{ and } v\}$$

and we define the map  $d : (A^*)^2 \rightarrow \mathbb{R}_+$  which sends any  $(u, v) \in (A^*)^2$  to  $2^{-r(u, v)}$ , with the conventions  $\min(\emptyset) = +\infty$  and  $2^{-\infty} = 0$ .

A profinite word is simply a Cauchy sequence for the profinite metric, up to equivalence of Cauchy sequences, defined as follows: two sequences of finite words  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  are equivalent if  $\lim_{n \rightarrow +\infty} d(x_n, y_n) = 0$ . For instance, every finite word  $u \in A^*$  can be seen as a profinite word, as it corresponds to the constant Cauchy sequence with the value  $u$ . It is relatively difficult to give concrete examples of profinite words which are not of this kind. One such example is what we denote by  $u^\omega$ : for any  $u \in A^*$ , one can prove that the sequence  $(u^{n!})_{n \in \mathbb{N}}$  is a Cauchy sequence (cf [?], Proposition 2.5), and thus has a limit

$$u^\omega := \lim_{n \rightarrow +\infty} u^{n!}.$$

- For any subset  $P$  of  $\mathbb{N}$ , we define the language of all words with length in  $P$ ,

$$L_P := \{w \in A^* : |w| \in P\} = |\cdot|^{-1}(P).$$

For any subsets  $P$  and  $Q$  of  $\mathbb{N}$ , we have

$$L_P \cup L_Q = \{w \in A^* : |w| \in P \text{ or } |w| \in Q\} = L_{P \cup Q},$$

$$L_P \cap L_Q = \{w \in A^* : |w| \in P \text{ and } |w| \in Q\} = L_{P \cap Q},$$

and

$$(L_P)^c = \{w \in A^* : |w| \notin P\} = L_{P^c}.$$

The set of languages  $\{L_P : P \subseteq \mathbb{N}\}$  forms a Boolean algebra that we denote by  $\mathcal{B}_{|\cdot|}$ . We observe that this Boolean algebra is isomorphic to  $\mathcal{P}(\mathbb{N})$ . Indeed, consider the map  $\varphi : \mathcal{B}_{|\cdot|} \rightarrow \mathcal{P}(\mathbb{N})$ , which sends any language of the form  $L_P$ , where  $P \subseteq \mathbb{N}$ ,

to  $P$ , and the map  $\psi : \mathcal{P}(\mathbb{N}) \rightarrow \mathcal{B}_{|\cdot|}$  which sends any  $P \subseteq \mathbb{N}$  to the language  $L_P$ . These two maps are Boolean homomorphisms and are each other's inverse. We conclude that  $\mathcal{S}(\mathcal{B}_{|\cdot|}) = \beta(\mathbb{N})$ .

- Fix a two letters alphabet  $A = \{a, b\}$ . Let us consider the language called *majority*, which corresponds to the set of all words which contain more occurrences of the letter  $a$  than the letter  $b$ ,

$$Maj := \{w \in A^* : |w|_a > |w|_b\}.$$

Considering a single language does not allow for an interesting treatment for topological recognition: for any language  $L \subseteq A^*$ , the Boolean algebra generated by  $L$  is isomorphic to the Boolean algebra  $\{\emptyset, L, L^c, A^*\}$ . Therefore, we consider a Boolean algebra that contains  $L$ , and is big enough to provide an interesting dual theoretic treatment: its closure under left and right quotients. We set

$$\mathcal{B}_{Maj} := \langle \{u^{-1}.Maj.v^{-1} : u, v \in A^*\} \rangle_{BA}.$$

One can prove (cf [28], Example 2.9) that  $\mathcal{B}_{Maj}$  is isomorphic to the Boolean sub-algebra of  $\mathcal{P}(\mathbb{Z})$

$$\mathcal{B}_{shift} := \langle \{\mathbb{Z}^+ - k : k \in \mathbb{Z}\} \rangle_{BA}$$

and that the dual space of this Boolean algebra is  $X := \mathbb{Z} \cup \{+\infty, -\infty\}$ , where

$$+\infty := \{K \in \mathcal{B}_{shift} : \text{the symmetric difference of } K \text{ and } \mathbb{Z}^+ \text{ is finite}\}$$

and

$$-\infty := \{K \in \mathcal{B}_{shift} : \text{the symmetric difference of } K \text{ and } \mathbb{Z}^- \text{ is finite,}\}$$

endowed with the following topology: a subset  $U$  of  $X$  is open if, and only if, it is contained in  $\mathbb{Z}$ ; or it contains  $+\infty$  and all but finitely many of the elements of  $\mathbb{Z}^+$ ; or it contains  $-\infty$  and all but finitely many of the elements of  $\mathbb{Z}^-$ .

## 1.4 Logic on words

In this section, we fix notations, and introduce the pieces of background required in order to follow Section 2.2, which treats the consequences of our study from the point of view of logic on words. The reader can refer to ([70], Chapter II) for a more complete account on logic on words, and to [72] and [1] for detailed illustrations of the interactions between logic and formal language theory.

Logic on words stems from the following idea: one way to think about a word  $w$  is as a relational structure over  $\{0, \dots, |w| - 1\}$ , equipped with a unary predicate  $a(\cdot)$ , for

every  $a \in A$ , which allows us to tell whether the letter at a given position of  $w$  is an  $a$ .

### 1.4.1 Syntax

We start by introducing the notion of numerical predicate, which will be the building blocks in order to define the formulas of logic on words. We then introduce the notion of uniformity, which, roughly speaking, will allow us to make the distinction between numerical predicates such that their interpretation takes into account the length of the words we will consider, and the ones that do not. In the next chapters, we mainly focus our study on uniform numerical predicates.

**Definition 1.41** (Numerical predicates). For any  $k \geq 0$ , a  $k$ -ary numerical predicate is a map

$$R^k : \mathbb{N}_{>0} \rightarrow \mathcal{P}(\mathbb{N}^k)$$

such that, for all  $n \geq 1$ ,  $R^k(n) \subseteq \{0, \dots, n-1\}^k$ . It is said to be *uniform* if there exists a subset  $Q \subseteq \mathbb{N}^k$  such that, for all  $n \geq 1$ ,

$$R^k(n) = Q \cap \{0, \dots, n-1\}^k.$$

We mention the particular case of 0-ary, or *nullary* predicates. According to the definition of numerical predicate we gave, a nullary predicate is a map  $\mathbb{N} \rightarrow \{0, 1\}$ , therefore it corresponds to considering a subset  $P$  of  $\mathbb{N}$ . In particular, a nullary predicate is necessarily uniform.

**Example 1.42.** We now give a few examples of numerical predicates.

- The unary numerical predicate

$$prime : \begin{cases} \mathbb{N}_{>0} & \longrightarrow \mathcal{P}(\mathbb{N}) \\ n & \longmapsto \{i \in \{0, \dots, n-1\} : i \text{ is prime} \} \end{cases}$$

is uniform, since for any  $n \geq 1$ ,  $prime(n) = \mathcal{P} \cap \{0, \dots, n-1\}$ , where  $\mathcal{P}$  is the set of all prime numbers.

- The binary numerical predicate

$$\leq : \begin{cases} \mathbb{N}_{>0} & \longrightarrow \mathcal{P}(\mathbb{N}^2) \\ n & \longmapsto \{(i, j) \in \{0, \dots, n-1\}^2 : i \leq j\} \end{cases}$$

is uniform, since for any  $n \geq 1$ ,  $\leq(n) = \mathcal{I} \cap \{0, \dots, n-1\}^2$ , where  $\mathcal{I}$  is the set of all of the couples of positive integers  $(i, j)$  such that  $i \leq j$ .

- The unary numerical predicate

$$\text{end} : \left| \begin{array}{ll} \mathbb{N}_{>0} & \longrightarrow \mathcal{P}(\mathbb{N}) \\ n & \longmapsto \{n-1\} \end{array} \right.$$

is a unary non-uniform numerical predicate. Indeed, there exists no subset  $P \subseteq \mathbb{N}$  such that, for all  $n \geq 1$ ,  $P \cap \{0, \dots, n-1\} = \{n-1\}$ .

Following the terminology introduced in [70], we now define the formulas of logic on words.

**Definition 1.43.** [Syntax of first-order logic on words.] We denote *first-order variables* by  $x, x_1, x_2$ , etc. We consider formulas that are recursively built from the following atomic blocks.

- *Letter predicates:* for every letter  $a \in A$ , a *letter predicate* is denoted by  $a(\cdot)$ . For any first-order variable  $x$ ,  $a(x)$  is an atomic formula.
- *Numerical predicates:* for any  $k \in \mathbb{N}$ , and for any list of  $k$  first-order variables  $x_1, \dots, x_k$ , if  $R^k : \mathbb{N}_{>0} \rightarrow \mathcal{P}(\mathbb{N}^k)$  is a  $k$ -ary numerical predicate, then  $R^k(x_1, \dots, x_k)$  is an atomic formula.

The closure operations on formulas are the following.

- If  $\varphi$  and  $\psi$  are formulas, then any Boolean combination of  $\varphi$  and  $\psi$  is a formula.
- If  $\varphi$  is a formula, and  $x$  is a variable, then  $\exists x\varphi(x)$  and  $\forall x\varphi(x)$  are formulas.

We say that a variable  $x$  occurs *freely* in a formula if it is not in the scope of a quantifier. In particular, we call *quantifier-free formulas* the Boolean combinations of atomic formulas. A *sentence* is a formula such that none of its variables are free. A *fragment of first-order logic* is a subset of the set of all sentences.

## 1.4.2 Semantics

As we previously announced, the particularity of logic on words is that we consider words as first-order structures. First, fix  $k \in \mathbb{N}$ . We use the notation  $\bar{i}$  to refer to the elements  $(i_1, \dots, i_k) \in \mathbb{N}^k$ , and  $\bar{a}$  to refer to the elements  $(a_1, \dots, a_k) \in A^k$ . For any finite word  $w \in A^*$ , we introduce the notation

$$|w|^k := \{\bar{i} \in \mathbb{N}^k : \forall j \in \{1, \dots, k\}, i_j < |w|\}.$$

Models of formulas with free variables among  $\bar{x} = \{x_1, \dots, x_k\}$ , where all the  $x_i$  are distinct, are given by elements  $(w, \bar{i})$  in

$$A^* \otimes \mathbb{N}^k := \{(w, \bar{i}) \in A^* \times \mathbb{N}^k : \bar{i} \in |w|^k\},$$



which we refer to as  $\bar{x}$ -structures, one marked position in the word corresponding exactly to one free variable. Note that several variables can mark the same position. It is important to emphasize that the notation  $\otimes$  used in this context is completely different from the notation used in Definition 1.27 in order to talk about product of closed subsets. We define an equivalence relation on formulas by saying that two formulas are equivalent if they have the same models.

We now introduce the semantic interpretation of the formulas we defined.

**Definition 1.44** (Semantics of logic on words). We define recursively the semantics of the formulas built in Definition 1.43. We start with the atomic formulas.

- For any  $l \in \{1, \dots, k\}$ , and for any letter  $a \in A$ , the  $\bar{x}$ -structure  $(w, \bar{i}) \in A^* \otimes \mathbb{N}^k$  satisfies  $a(x_{i_l})$  if, and only, if  $w_{i_l} = a$ . In particular, for any  $\bar{a} \in A^k$ , we use the notation  $w[\bar{i}] = \bar{a}$  in order to say that the  $\bar{x}$ -structure satisfies the formula

$$\bigwedge_{j=1}^k a_j(x_j),$$

that is, for every  $j \in \{1, \dots, k\}$ ,  $w_{i_j} = a_j$ .

- For any  $l \in \{1, \dots, k\}$ , any  $j_1, \dots, j_l \leq k$ , and any  $l$ -ary numerical predicate  $R^l$ , the  $\bar{x}$ -structure  $(w, \bar{i}) \in A^* \otimes \mathbb{N}^k$  satisfies  $R^l(x_{j_1}, \dots, x_{j_l})$  if, and only, if  $(i_{j_1}, \dots, i_{j_l})$  belongs to  $R^l(|w| - 1)$ .

Now, the closure operations on formulas are defined as follows.

- The Boolean operations are interpreted in the usual way.
- For any  $l \in \{1, \dots, k\}$ , and any  $i_{\setminus l} := (i_j)_{\substack{1 \leq j \leq k \\ j \neq l}} \in \mathbb{N}^{k-1}$ , given a formula  $\varphi(\bar{x})$ , a  $(\bar{x} \setminus \{x_l\})$ -structure  $(w, \bar{i}_{\setminus l}) \in A^* \otimes \mathbb{N}^{k-1}$  satisfies the formula  $\exists x_l \varphi(\bar{x})$  if, and only, if there exist  $i_l < |w|$  such that  $(w, \bar{i}) \in A^* \otimes \mathbb{N}^k$  satisfies  $\varphi(\bar{x})$ .

Note that the interpretation of nullary predicates  $P \subseteq \mathbb{N}$  on words is the following: the  $\emptyset$ -structure  $w$  satisfies  $P$  if, and only if,  $|w| \in P$ .

*Remark 1.45.* In particular, if  $R^k$  is a  $k$ -ary uniform numerical predicate, since there exists a subset  $Q \subseteq \mathbb{N}^k$  such that, for all  $n \geq 1$ ,  $R^k(n) = Q \cap \{0, \dots, n-1\}^k$ , a  $\bar{x}$ -structure  $(w, \bar{i}) \in A^* \otimes \mathbb{N}^k$  satisfies  $R^k(\bar{x})$  if, and only if,  $\bar{i}$  belongs to  $Q$ .

**Example 1.46.** We introduced in Example 1.42 the uniform binary numerical predicate  $\leq$ . For any variables  $x, y$ , and any letters  $a, b \in A$ , the quantifier-free formula

$$\varphi(x, y) = a(x) \wedge b(y) \wedge \leq(x, y)$$

corresponds to the set  $L_{\varphi(x, y)}$  of all elements  $(w, i, j)$  in  $A^* \otimes \mathbb{N}^2$  such that  $w_i = a, w_j = b$  and  $i \leq j$ ,

$$L_{\varphi(x, y)} = \{(w, i, j) \in A^* \otimes \mathbb{N}^2 : w[i, j] = (a, b) \text{ and } (i, j) \in \mathcal{I}\}.$$

Applying a layer of existential quantifiers to this formula leads us to consider the sentence

$$\exists x \exists y \varphi(x, y),$$

which corresponds to the language consisting in all words  $w \in A^*$  such that there exist two positions  $i, j < |w|$  such that  $w_i = a, w_j = b$  and  $i \leq j$ , in other words, the language  $A^*aA^*bA^*$ .

For any formula  $\varphi$  on the set of variables  $\bar{x} = \{x_1, \dots, x_k\}$ , and for any  $\bar{x}$ -structures  $(w, \bar{i})$ , we use the notation

$$(w, \bar{i}) \models \varphi(\bar{x})$$

to say that  $(w, \bar{i})$  satisfies the formula  $\varphi$ . We denote by  $L_\varphi$  the subset of  $A^* \otimes \mathbb{N}^k$  of all the  $\bar{x}$ -structures  $(w, \bar{i})$  satisfying the formula  $\varphi$ ,

$$L_\varphi := \{(w, \bar{i}) \in A^* \otimes \mathbb{N}^k : (w, \bar{i}) \models \varphi(\bar{x})\}.$$

Notice that  $L_\varphi$  is a language of finite words on the alphabet  $A$  if, and only, if the formula  $\varphi$  is a sentence. Whenever  $F$  is a subset of the set of formulas with free variables among  $\bar{x}$ , which is closed under the Boolean connectives, the collection  $\{L_\varphi : \varphi \in F\}$  is a Boolean subalgebra of  $\mathcal{P}(A^* \otimes \mathbb{N}^k)$ . This simple observation is what allows us to apply duality for Boolean algebras to logic fragments, and motivates the approach taken in Chapter 2.

For any  $k \geq 1$ , we denote by

$$FO[\mathcal{N}_k^u]$$

the set of languages corresponding to the fragment of first order logic defined by Boolean combinations of first-order sentences built by only using letter predicates, and uniform numerical predicates with arity  $l \in \{1, \dots, k\}$ . If we want to add nullary predicates, then we use the notation  $FO[\mathcal{N}_0, \mathcal{N}_k^u]$ . Finally, we use the notation  $FO[\mathcal{N}_k]$  if we consider  $k$ -ary numerical predicates that are not necessarily uniform.

## 1.5 Ultrafilter equations

As we announced in the general introduction, the foundation of descriptive complexity theory is that it is possible to express computational complexity classes in terms of fragments of logic on words. In Section 1.4, we explained how one can describe certain fragments of first-order logic in terms of Boolean subalgebras of  $\mathcal{P}(A^*)$ . Fixing two such Boolean algebras  $\mathcal{B}_1, \mathcal{B}_2$ , proving that they are distinct amounts to constructing a language  $L \subseteq A^*$  which belongs to  $\mathcal{B}_1$ , but not to  $\mathcal{B}_2$ . Let us take the point of view of duality theory. The dual space  $\mathcal{S}(\mathcal{B})$  is a topological object canonically associated to  $\mathcal{B}$ , however, in general, it happens to be too “big” to constitute a practical description of  $\mathcal{B}$ . A question that arises naturally is therefore the following: is it possible to intro-

duce a practical topological object which holds enough information to characterize the Boolean algebra we are interested in? The answer involves introducing the notion of *ultrafilter equation*, a well-chosen family of pairs of ultrafilters in the dual space. The main ingredient we need to use is the fact that any Boolean algebra of languages is a subalgebra of  $\mathcal{P}(A^*)$ . A Boolean algebra of languages  $\mathcal{B}$  is given by an embedding  $\mathcal{B} \hookrightarrow \mathcal{P}(A^*)$  and thus the dual map  $\beta(A^*) \rightarrow \mathcal{S}(\mathcal{B})$  is a quotient, given by equating elements in the dual space. The idea would be to find families of pairs of points in the dual space (ideally, much smaller than  $\mathcal{S}(\mathcal{B})^2$ ) that allow for a characterisation of  $\mathcal{B}$ . Since the reasoning applies for any subalgebra of a given Boolean algebra we formalize the reasoning in this setting.

**Definition 1.47.** For any Boolean space  $X$ , a *Boolean equivalence relation* is an equivalence relation  $\mathcal{E}$  of  $X$  such that the quotient space  $X/\mathcal{E}$  is also a Boolean space.

**Definition 1.48.** For any Boolean algebra  $\mathcal{B}$ , any two ultrafilters  $\gamma_1, \gamma_2 \in \mathcal{S}(\mathcal{B})$ , and any  $b \in \mathcal{B}$ , we say that  $b$  satisfies the  *$\mathcal{B}$ -equation*  $\gamma_1 \leftrightarrow \gamma_2$  if, and only if,

$$b \in \gamma_1 \iff b \in \gamma_2.$$

**Theorem 1.49** (Stone duality for Boolean subalgebras, [65], Theorem 5.1). *Let  $\mathcal{B}$  be a Boolean algebra, and  $X$  its associated dual space. Let us consider the map from  $\mathcal{P}(\mathcal{B})$  to  $\mathcal{P}(X^2)$  which sends any subset  $S$  of  $\mathcal{B}$  to*

$$\{(x, y) \in X : \forall b \in S, (b \in x \iff b \in y)\}$$

*and the map from  $\mathcal{P}(X^2)$  to  $\mathcal{P}(\mathcal{B})$  which sends any subset  $E$  of  $X^2$  to*

$$\{b \in \mathcal{B} : \forall (x, y) \in E, (b \in x \iff b \in y)\}.$$

*These maps establish a Galois connection whose Galois closed sets are the Boolean equivalence relations on  $X$  and the Boolean subalgebras of  $\mathcal{B}$  respectively. In particular, every set of equations over  $X$  determines a Boolean subalgebra of  $\mathcal{B}$ , and every Boolean subalgebra of  $\mathcal{B}$  is given by a set of equations over  $X$ .*

**Corollary 1.50.** *Any Boolean algebra of languages on a finite alphabet  $A$  can be defined by a set of equations of the form  $\gamma_1 \leftrightarrow \gamma_2$  where  $\gamma_1$  and  $\gamma_2$  are ultrafilters on the set of words.*

If the Boolean algebra we consider in Theorem 1.49 is closed under quotients, then the set of all equations satisfied by it is a kind of congruence. For any ultrafilter on words  $\gamma \in \beta(A^*)$  and any word  $w \in A^*$ , set

$$w.\gamma := \{w^{-1}.L : L \in \gamma\} \text{ and } \gamma.w := \{L.w^{-1} : L \in \gamma\}.$$

Fix two ultrafilters  $\gamma_1$  and  $\gamma_2$  in  $\beta(A^*)$ . It is a simple computation to check that, if a Boolean algebra closed under quotients satisfies the equation  $\gamma_1 \leftrightarrow \gamma_2$ , then it also satisfies the equations  $w.\gamma_1 \leftrightarrow w.\gamma_2$  and  $\gamma_1.w \leftrightarrow \gamma_2.w$ , for any word  $w \in A^*$ . In view of this result, it is convenient to introduce the following notation. Given two ultrafilters  $\gamma_1$  and  $\gamma_2$  in  $\beta(A^*)$ , we say that a language satisfies the ultrafilter equation  $\gamma_1 = \gamma_2$  if it satisfies all the ultrafilter equations  $w.\gamma_1 \leftrightarrow w.\gamma_2$  and  $\gamma_1.w \leftrightarrow \gamma_2.w$ , for all words  $w \in A^*$ .

In the particular case of Boolean algebras of regular languages, an even stronger result is available. As we explained in Example 1.40, the dual space of  $\text{Reg}(A^*)$  is  $\widehat{A^*}$ , the set of all profinite words on  $A^*$ . Theorem 1.49 applied to  $\text{Reg}(A^*)$  shows that any Boolean algebra of regular languages can be defined by a set of equations of the form  $u \leftrightarrow v$ , where  $u$  and  $v$  are profinite words.

**Theorem 1.51** ([29], Proposition 1.3). *Let  $\mathcal{B}$  be a Boolean algebra of regular languages of  $A^*$  closed under quotients and let  $w_1, w_2 \in \widehat{A^*}$ . If  $\mathcal{B}$  satisfies the profinite equation  $w_1 \leftrightarrow w_2$ , then it satisfies the profinite equations  $u.w_1 \leftrightarrow u.w_2$  and  $w_1.u \leftrightarrow w_2.u$ , for each profinite word  $u \in \widehat{A^*}$ .*

Given  $w_1, w_2 \in \widehat{A^*}$ , we say that a regular language satisfies the profinite equation  $w_1 = w_2$  if it satisfies the profinite equations  $u.w_1 \leftrightarrow u.w_2$  and  $w_1.u \leftrightarrow w_2.u$ , for each profinite word  $u \in \widehat{A^*}$ . The main interest of these notations is to allow one to produce smaller sets of defining equations for Boolean algebra of languages which are closed under quotients.

**Example 1.52.** We list a few examples of profinite equations corresponding to fragments of first-order logic.

- The fragment  $FO[<]$  of first-order logic defined as the set of all sentences written by only using letter predicates and the binary numerical predicate  $<$  introduced in Example 1.42 is described by the profinite equations

$$u^\omega = u^{\omega+1},$$

where  $u \in A^*$ .

- The fragment  $\mathcal{B}\Sigma_1[<]$  of first-order logic consisting in Boolean combinations of sentences written by using the binary numerical predicate  $<$ , and letter predicates is described by the profinite equations

$$(uv)^\omega = (vu)^\omega \text{ and } u^\omega = u^{\omega+1},$$

where  $u, v \in A^*$ .

- The fragment  $FO_2[<]$  first-order logic defined as the set of all sentences, written by only using letter predicates and the binary numerical predicate  $<$  and exactly

two first-variables is described by the profinite equations

$$(uv)^\omega(vu)^\omega(uv)^\omega = (uv)^\omega \text{ and } u^\omega = u^{\omega+1},$$

where  $u, v \in A^*$ .

The reader can refer to [51], [53] [12] and [59] for more fragments of logic corresponding to families of regular languages, and their description in terms of profinite equations. We conclude the section by giving an example of ultrafilter equations for a Boolean algebra of non-regular languages.

**Example 1.53.** The Boolean algebra  $\mathcal{B}_{Maj}$  that we introduced in Example 1.40 admits the basis of ultrafilter equations

$$\mu = \mu + 1,$$

for every  $\mu \in {}^*(\mathbb{Z})$ , where

$$\mu + 1 := \{P - 1 : P \in \mu\},$$

see [28], Example 2.21 for a proof.

---

## Duality for the existential fragment of first-order logic on words

---

This thesis lies at the intersection of formal language theory and duality theoretic methods. The contribution of this chapter is a deepening of the knowledge currently available on existential quantification for logic on words. More precisely, we provide a characterisation of the dual space of the Boolean algebra corresponding to the first-order sentences built by using uniform numerical predicates of arity  $k$ , for a fixed  $k \in \mathbb{N}$  (Theorem 2.14). The consequences of applying one layer of existential quantifier to Boolean algebras of languages defined by formulas with free first-order variables, and their counterpart at the level of topological recognisers have already been well studied in [30]. On the algebraic side, we apply one layer of existential quantifier, while on the topological side, we take the Vietoris hyperspace. Yet, unlike in the case of finite recognisers where a minimisation algorithm is available, there exists, at the moment, no well-known procedure which would allow to directly derive the minimal topological recognizer out of a given topological recogniser. Indeed, the minimal topological recognizer corresponds to the dual space of the Boolean algebra in question, and very few concrete computations of dual spaces for fragments of logic on words which lie outside of the regular case are available. This work provides a thorough study in the case where the Boolean algebra we quantify over consists of exactly every quantifier-free formula, and we add one layer of quantification.

Apart from being an object of study in itself, another reason which motivates our interest in this question is that its answer has an application in computer science, more precisely in complexity theory. A result of Immerman [36] establishes strong connections between computational complexity classes and formal language theory: it states, in particular, that we can associate to most complexity classes a class of formal languages. This is at the foundation of what is now called descriptive complexity theory. Therefore, the problem of separating classes in complexity theory amounts to separating the

corresponding class of languages, and this is where duality theory comes into play. Let us give more details for the case we focus on in this thesis. Fix a finite alphabet  $A$ . On the complexity theory side,  $AC0$  is a complexity class defined by sequences of Boolean circuits. It is important to note that, in this thesis, we do not define Boolean circuits, neither do we work with them: we only work with their logical counterpart, see [70] for a detailed introduction to the topic. The corresponding class of languages associated to  $AC0$  is

$$FO[\mathcal{N}] := \bigcup_{k \in \mathbb{N}} FO[\mathcal{N}_k],$$

see [70], Theorem IX.2.1, where, for every  $k \in \mathbb{N}$ ,  $FO[\mathcal{N}_k]$  has been defined at the end of Section 1.4. An approach to describing  $FO[\mathcal{N}]$  would be to start by describing smaller fragments, and build on them to eventually characterise this class of languages. For each letter  $a$  in  $A$ , and for each subset  $P$  of  $\mathbb{N}$ , let us introduce the  $a$ -content of  $w$ ,

$$c_a(w) := \{i \in |w| : w_i = a\},$$

and the languages

$$L_P := \{w \in A^* : |w| \in P\}$$

and

$$L_P^a := \{w \in A^* : c_a(w) \cap P \neq \emptyset\}.$$

It has been proven in [29] that a language  $L$  belongs to the Boolean algebra generated by the languages  $L_P$  and  $L_P^a$ , where  $a$  ranges over  $A$ , and  $P$  ranges over the subsets of  $\mathbb{N}$ , if, and only if,  $L$  belongs to the class of languages defined by first-order sentences built on nullary predicates and unary uniform numerical predicates, that is  $FO[\mathcal{N}_0, \mathcal{N}_1^u]$ . A characterisation of the dual space of this Boolean algebra had already been discovered in an unpublished paper of Gehrke, Krebs and Pin. This allowed for a description of this fragment of first-order logic on words in [29] in terms of ultrafilter equations, that is, a family of pairs of ultrafilter on words which are sufficient to characterise it. The idea is the following: a Boolean algebra of languages  $\mathcal{B}$  is a subalgebra of  $\mathcal{P}(A^*)$ , and thus the canonical embedding provides a continuous quotient map  $\pi : \beta(A^*) \twoheadrightarrow \mathcal{S}(\mathcal{B})$ . By definition,  $\pi$  sends an ultrafilter  $\gamma \in \beta(A^*)$  to  $\{L \in \mathcal{B} : L \in \gamma\}$ . Therefore, an equivalent way to say that  $\mathcal{B}$  satisfies the ultrafilter equation  $\gamma_1 \leftrightarrow \gamma_2$ , for a pair  $(\gamma_1, \gamma_2) \in \beta(A^*)^2$ , is to say that  $\pi(\gamma_1) = \pi(\gamma_2)$ .

**Outline of the chapter:** Our wish is to get an understanding of a larger fragment of  $FO[\mathcal{N}]$  than  $FO[\mathcal{N}_1^u]$ , the class of languages defined by first-order sentences built on unary uniform numerical predicates. We denote this larger fragment by  $\mathcal{B}_k$ . It would take into account the uniform numerical predicates of arity lesser or equal to  $k$ , for some  $k \geq 1$ , but it would not take into account alternation of quantifiers. In section 2.1, we set the notations and introduce the Boolean algebra  $\mathcal{B}_k$ , for any  $k \geq 1$ , and its dual space  $X_k$ , that will be at the center of our study in this chapter. In section 2.2, we

give a description of  $\mathcal{B}_k$  from the perspective of logic on words, and in section 2.3 we give a first characterisation of the dual space  $X_k$  in terms of finite colourings of  $\mathbb{N}^k$ . This characterisation being a bit abstract, we study in section 2.4 the dual space of a subalgebra of  $\mathcal{B}_k$  that is big enough to give an insight on how to obtain a more concrete characterisation of  $X_k$ .

## 2.1 General setting

Fix a finite alphabet  $A$ . We introduce for any  $k \geq 1$ , and any  $k$ -tuple of letters  $\bar{a} \in A^k$  the map  $c_{\bar{a}} : A^* \rightarrow \mathcal{P}_{fin}(\mathbb{N}^k)$ , where  $\mathcal{P}_{fin}(\mathbb{N}^k)$  is the set of all finite subsets of  $\mathbb{N}^k$ , which sends a finite word  $w$  to its  $\bar{a}$ -content,

$$c_{\bar{a}}(w) := \{\bar{i} \in |\mathbb{N}^k| : w[\bar{i}] = \bar{a}\}.$$

**Definition 2.1.** For any subset  $Q$  of  $\mathbb{N}^k$ , we introduce the languages

$$L_{\Diamond_{\bar{a}} Q} := \{w \in A^* : c_{\bar{a}}(w) \cap Q \neq \emptyset\}$$

and

$$L_{\Box_{\bar{a}} Q} := (L_{\Diamond_{\bar{a}} Q^c})^c = \{w \in A^* : c_{\bar{a}}(w) \subseteq Q\}$$

This notation allows us to keep the intuition of modal algebra, as introduced in section 1.2.

**Example 2.2.** We provide a few concrete examples of these languages in the case where  $k = 2$ , which shall be used in order to build an intuition over  $\mathcal{B}_k$ . Fix two letters  $a$  and  $b$  in  $A$ .

- Assume that  $Q$  is a subset of  $\mathbb{N}^2$  such that there exist two subsets  $P$  and  $P'$  of  $\mathbb{N}$  such that  $Q = P \times P'$ . In this case, we have that

$$\begin{aligned} L_{\Diamond_{P \times P'}^{a,b}} &= \{w \in A^* \mid c_{a,b}(w) \cap (P \times P') \neq \emptyset\} \\ &= \{w \in A^* \mid (c_a(w) \times c_b(w)) \cap (P \times P') \neq \emptyset\} \\ &= \{w \in A^* \mid c_a(w) \cap P \neq \emptyset\} \cap \{w \in A^* \mid c_b(w) \cap P' \neq \emptyset\} \\ &= L_{\Diamond_P^a} \cap L_{\Diamond_{P'}^b}. \end{aligned}$$

- Assume that  $Q$  is equal to  $\Delta$ , the diagonal of  $\mathbb{N}^2$ , in other terms every pair  $(i, i)$  with  $i \in \mathbb{N}$ . Then we have

$$\begin{aligned} L_{\Diamond_{\Delta}^{a,a}} &= \{w \in A^* : c_{a,a}(w) \cap \Delta \neq \emptyset\} \\ &= \{w \in A^* : \exists i \in \mathbb{N}, w_i = a\} \\ &= A^* a A^*. \end{aligned}$$



- Assume that  $Q$  is the subset  $\{(i, j) \in \mathbb{N}^2 \mid i \leq j\}$ , a similar reasoning allows us to prove that

$$L_{\diamond_{\{(i,j) \in \mathbb{N}^2 \mid i \leq j\}}}^{a,a} = L_{\diamond_{\Delta}}^{a,a}.$$

- Assume that  $Q$  is the subset  $\{(i, i+1) \mid i \in \mathbb{N}\}$ . Then we have

$$\begin{aligned} L_{\diamond_{\{(i,i+1) \mid i \in \mathbb{N}\}}}^{a,a} &= \{w \in A^* : c_{a,a}(w) \cap \{(i, i+1) \mid i \in \mathbb{N}\} \neq \emptyset\} \\ &= \{w \in A^* : \exists i \in \mathbb{N}, w[(i, i+1)] = (a, a)\} \\ &= A^*aaA^*. \end{aligned}$$

Let  $\mathcal{B}_k$  be the Boolean subalgebra of  $\mathcal{P}(A^*)$  generated by the languages  $L_{\diamond_{\bar{a}}}^{a,a}$ , where  $\bar{a}$  ranges over  $A^k$ , and  $Q$  ranges over the subsets of  $\mathbb{N}^k$ ,

$$\mathcal{B}_k := \langle \{L_{\diamond_{\bar{a}}}^{a,a} : \bar{a} \in A^k, Q \subseteq \mathbb{N}^k\} \rangle_{BA}.$$

The main purpose of this chapter is to understand, and give characterisations of the dual space of the Boolean algebra  $\mathcal{B}_k$ . We denote by  $X_k$  the dual space of  $\mathcal{B}_k$ .

We denote by  $V_k$  the  $A^k$ -fold power of Vietoris hyperspaces

$$V_k := \mathcal{V}(\beta(\mathbb{N}^k))^{A^k}.$$

We consider the function

$$c^k : A^* \rightarrow V_k,$$

which sends a finite word  $w \in A^*$  to the following family of clopen subsets,

$$(\widehat{c_{\bar{a}}(w)})_{\bar{a} \in A^k}.$$

By the universal property of Čech–Stone compactification, there exists a unique continuous map  $c^k : \beta(A^*) \rightarrow V_k$  which extends it. We give a description of the image of this map: it actually corresponds to the dual space of  $\mathcal{B}_k$ .

**Proposition 2.3.** *The image of  $c^k$  is homeomorphic to  $X_k$ .*

*Proof.* Let us denote by  $\mathcal{M}_k$  the  $A^k$ -fold copower of  $M\mathcal{P}(\mathbb{N}^k)$ , that is the free Boolean algebra generated by the formal generators  $\diamond_{\bar{a}}Q$ , where  $\bar{a}$  ranges over  $A^k$ , and  $Q$  ranges over the subsets of  $\mathbb{N}^k$ . As we mentioned in Proposition 1.30, the dual of  $M\mathcal{P}(\mathbb{N}^k)$  is  $\mathcal{V}(\beta(\mathbb{N}^k))$ . As we observed in Proposition 1.13, Boolean Stone duality turns coproducts into products: the dual space of the Boolean algebra  $\mathcal{M}_k$  is  $V_k$ . We start by defining a Boolean algebra homomorphism  $h_k : \mathcal{M}_k \rightarrow \mathcal{P}(A^*)$  which we will prove to be dual to the continuous map  $c^k : \beta(A^*) \rightarrow V_k$ . First, fixing  $\bar{a} \in A^k$ , we consider the map

$$L_{\diamond_{(\cdot)}}^{\bar{a}} : \mathcal{P}(\mathbb{N}^k) \rightarrow \mathcal{P}(A^*)$$

which sends any  $Q \subseteq \mathbb{N}^k$  to the language  $L_{\diamond_{\bar{a}} Q}$ . This map preserves finite joins: indeed,  $L_{\diamond_{\emptyset} \bar{a}} = \emptyset$  and, for any  $Q_1, Q_2 \subseteq \mathbb{N}^k$ ,

$$\begin{aligned} L_{\diamond_{\bar{a}} Q_1} \cup L_{\diamond_{\bar{a}} Q_2} &= \{w \in A^* : c_{\bar{a}}(w) \cap Q_1 \neq \emptyset\} \cup \{w \in A^* : c_{\bar{a}}(w) \cap Q_2 \neq \emptyset\} \\ &= \{w \in A^* : c_{\bar{a}}(w) \cap Q_1 \neq \emptyset \text{ or } c_{\bar{a}}(w) \cap Q_2 \neq \emptyset\} \\ &= \{w \in A^* : c_{\bar{a}}(w) \cap (Q_1 \cup Q_2) \neq \emptyset\} \\ &= L_{\diamond_{\bar{a}} (Q_1 \cup Q_2)}. \end{aligned}$$

Therefore, this join preserving map extends uniquely to a Boolean algebra homomorphism

$$h^{\bar{a}} : MP(\mathbb{N}^k) \rightarrow \mathcal{P}(A^*).$$

We now define  $h_k : \mathcal{M}_k \rightarrow \mathcal{P}(A^*)$  by using the universal property of the  $A^k$ -fold copower of  $MP(\mathbb{N}^k)$ , that is  $h_k$  is the unique Boolean algebra homomorphism such that, for any  $\bar{a} \in A^k$ , and for any  $Q \subseteq \mathbb{N}^k$ ,

$$h_k(\diamond_{\bar{a}} Q) = L_{\diamond_{\bar{a}} Q}.$$

In particular, this equality proves that  $Im(h_k) = \mathcal{B}_k$ , since  $\mathcal{M}_k$  is the Boolean algebra generated by the elements of the form  $\diamond_{\bar{a}} Q$ , where  $\bar{a}$  ranges over  $A^k$ , and  $Q$  ranges over the subsets of  $\mathbb{N}^k$ . Therefore, we have the following commutative diagram in **Bool**.

$$\begin{array}{ccc} \mathcal{M}_k & \xrightarrow{h_k} & \mathcal{P}(A^*) \\ & \searrow & \uparrow \\ & & \mathcal{B}_k \end{array}$$

Now, by duality, we have the following diagram in **Stone**.

$$\begin{array}{ccc} \beta(A^*) & \xrightarrow{(h_k)^{-1}} & V_k \\ & \searrow & \uparrow \\ & & X_k \end{array}$$

In order to conclude that  $Im(c^k) = X_k$ , it is enough to prove that  $c^k$  is dual to  $h_k$ , that is that  $c^k = (h_k)^{-1}$ . Since we consider continuous maps between compact Hausdorff spaces, and  $A^*$  is a dense subspace of  $\beta(A^*)$ , we only need to prove that the restriction of these maps to  $A^*$  are equal. Now, for any word  $w \in A^*$ , we have by definition

$$c^k(w) = (c_{\bar{a}}(w))_{\bar{a} \in A^k},$$

and since the duality turns coproducts into products,

$$(h_k)^{-1}(w) = ((h^{\bar{a}})^{-1}(w))_{\bar{a} \in A^k}.$$

Finally, for every  $w \in A^*$ , and every  $\bar{a} \in A^k$ ,

$$\begin{aligned}
(h^{\bar{a}})^{-1}(w) &= \{\bar{i} \in \mathbb{N}^k : w \in h^{\bar{a}}(\diamond^{\bar{a}}(\{\bar{i}\}))\} \\
&= \{\bar{i} \in \mathbb{N}^k : w \in L_{\diamond^{\bar{a}}_{\{\bar{i}\}}}^{\bar{a}}\} \\
&= \{\bar{i} \in \mathbb{N}^k : c_{\bar{a}}(w) \cap \{\bar{i}\} \neq \emptyset\} \\
&= \{\bar{i} \in \mathbb{N}^k : w[\bar{i}] = \bar{a}\} \\
&= c_{\bar{a}}(w),
\end{aligned}$$

and we conclude that  $c^k = (h_k)^{-1}$ , and therefore that  $Im(c^k) = X_k$ .  $\square$

## 2.2 Logical description of $\mathcal{B}_k$

Fix  $k \geq 1$ . In this section, we give a description of the Boolean algebra  $\mathcal{B}_k$  in the context of logic on words. The Boolean algebra  $\mathcal{B}_k$  contains the subalgebra generated by the languages  $L_P^a$ , where  $a$  ranges over  $A$ , and  $P$  ranges over the subsets of  $\mathbb{N}$ . This implies that it already encodes at the very least all of the first-order sentences built by using unary uniform numerical predicates (see [29], Theorem 2.9 for a proof). We will prove that  $\mathcal{B}_k$  can actually be identified with the Boolean algebra of languages corresponding to formulas which are Boolean combinations of sentences defined by using a block of at most  $k$  existential quantifiers, letter predicates and uniform numerical predicates of arity  $l \in \{1, \dots, k\}$ .

Fixing a set of free variables  $\bar{x} = \{x_1, \dots, x_k\}$ , we first express any quantifier-free formula written by using only a subset of these variables as a normal form which involves exactly all of the free-variables in  $\bar{x}$ .

**Lemma 2.4.** *Any quantifier-free formula  $\varphi$  such that the set of its variables  $\{x_{j_1}, \dots, x_{j_l}\}$  is a subset of  $\bar{x}$  can be written as a formula of the form*

$$\bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge R^{\bar{a}}(\bar{x})),$$

where, for every  $\bar{a} \in A^k$ ,  $\bar{a}(\bar{x})$  is defined as the conjunction  $\bigwedge_{j=1}^k a_j(x_j)$ , and  $R^{\bar{a}}$  is a  $k$ -ary numerical predicate.

*Proof.* We denote by  $\iff$  the relation of logical equivalence between formulas. We prove the statement by structural induction on quantifier-free formulas. We start with atomic formulas. For letter predicates, we have that, for any  $j \in \{1, \dots, k\}$ , and any  $a \in A$ ,

$$a(x_j) \iff \bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge S^{\bar{a}}(\bar{x})),$$

where, for every  $\bar{a} \in A^k$ ,  $S^{\bar{a}}$  is the  $k$ -ary numerical predicate defined as

$$S^{\bar{a}} : \begin{cases} \mathbb{N}_{>0} & \longrightarrow \mathcal{P}(\mathbb{N}^k) \\ n & \longmapsto \begin{cases} \{0, \dots, n-1\}^k & \text{if } a_j = a \\ \emptyset & \text{otherwise} \end{cases} \end{cases}.$$

For numerical predicates, we have that, for any  $l$ -ary numerical predicate  $R^l$ ,

$$R^l(x_{j_1}, \dots, x_{j_l}) \iff \bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge T_{R^l}^{\bar{a}}(\bar{x})),$$

where, for every  $\bar{a} \in A^k$ ,  $T_{R^l}^{\bar{a}}$  is  $T_{R^l}$ , the  $k$ -ary numerical predicate defined as

$$T_{R^l} : \begin{cases} \mathbb{N}_{>0} & \longrightarrow \mathcal{P}(\mathbb{N}^k) \\ n & \longmapsto \{(n_1, \dots, n_k) \in \{0, \dots, n-1\}^k \mid (n_{j_1}, \dots, n_{j_l}) \in R^l(n)\} \end{cases}.$$

To conclude, all we need to do is to prove that formulas of the form  $\bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge R^{\bar{a}}(\bar{x}))$  are closed under Boolean operations. For any formulas  $\varphi_1(\bar{x})$  of the form  $\bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge R_1^{\bar{a}}(\bar{x}))$  and  $\varphi_2(\bar{x})$  of the form  $\bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge R_2^{\bar{a}}(\bar{x}))$ , we have that

$$\varphi_1(\bar{x}) \wedge \varphi_2(\bar{x}) \iff \bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge (R_1^{\bar{a}} \cap R_2^{\bar{a}})(\bar{x}))$$

where, for every  $\bar{a} \in A^k$ ,  $(R_1^{\bar{a}} \cap R_2^{\bar{a}})$  is the  $k$ -ary numerical predicate defined as

$$R_1^{\bar{a}} \cap R_2^{\bar{a}} : \begin{cases} \mathbb{N}_{>0} & \longrightarrow \mathcal{P}(\mathbb{N}^k) \\ n & \longmapsto R_1^{\bar{a}}(n) \cap R_2^{\bar{a}}(n) \end{cases},$$

and

$$\varphi(\bar{x}) \vee \psi(\bar{x}) \iff \bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge (R_1^{\bar{a}} \cup R_2^{\bar{a}})(\bar{x})).$$

where, for every  $\bar{a} \in A^k$ ,  $(R_1^{\bar{a}} \cup R_2^{\bar{a}})$  is defined in an analogous way. Finally, the negation of any atomic formula is equivalent to a disjunction of atomic formulas: for any  $l$ -ary predicate  $R^l$ , and any free-variables  $(x_{j_1}, \dots, x_{j_l})$ ,

$$\neg R^l(x_{j_1}, \dots, x_{j_l}) \iff ((R^l)^c)(x_{j_1}, \dots, x_{j_l}),$$

where, for every  $\bar{a} \in A^k$ ,  $(R^l)^c$  is the  $k$ -ary numerical predicate defined as

$$(R^l)^c : \begin{cases} \mathbb{N}_{>0} & \longrightarrow \mathcal{P}(\mathbb{N}^k) \\ n & \longmapsto \{0, \dots, n-1\}^k \setminus T_{R^l}(n) \end{cases};$$

and for any letter  $a$  and any  $j \in \{1, \dots, k\}$ ,

$$\neg a(x_j) \iff \bigvee_{\substack{b \in A \\ b \neq a}} b(x_j).$$

□

Therefore, by Lemma 2.4, if we apply one layer of existential quantifier to a quantifier-free formula  $\varphi$  whose set of variables is contained in  $\bar{x}$ , we obtain a sentence of the form

$$\exists x_1 \dots \exists x_k \bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge R^{\bar{a}}(\bar{x}))$$

where, for every  $\bar{a} \in A^k$ ,  $R^{\bar{a}}$  is a  $k$ -ary numerical predicate.

From now on, we will restrict our attention to uniform numerical predicates. By Remark 1.45, this means that, for every  $\bar{a} \in A^k$ , considering the  $k$ -ary numerical predicate  $R^{\bar{a}}$  is equivalent to considering a subset  $Q^{\bar{a}} \subseteq \mathbb{N}^k$ . For every  $A^k$ -indexed family  $\mathcal{Q} = (Q^{\bar{a}})_{\bar{a} \in A^k}$  of  $k$ -ary numerical predicates, you denote by  $\psi(\mathcal{Q})$  the formula

$$\psi(\mathcal{Q}) := \exists x_1 \dots \exists x_k \bigvee_{\bar{a} \in A^k} (\bar{a}(\bar{x}) \wedge Q^{\bar{a}}(\bar{x})).$$

We define  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ , the Boolean algebra generated by the languages  $\psi(\mathcal{Q})$ , where  $\mathcal{Q}$  ranges over all such families.

Since the existential quantifier commutes with finite disjunctions, this Boolean algebra is generated by the languages corresponding to sentences of the form

$$\bigvee_{\bar{a} \in A^k} \exists x_1 \dots \exists x_k \bar{a}(\bar{x}) \wedge Q^{\bar{a}}(\bar{x}).$$

Now, note that, for any  $\bar{a} \in A^k$ , and any  $Q^{\bar{a}} \subseteq \mathbb{N}^k$ , the sentence

$$\exists x_1 \dots \exists x_k \bar{a}(\bar{x}) \wedge Q^{\bar{a}}(\bar{x})$$

corresponds to the language of the form

$$\{w \in A^* : \exists \bar{i} \in Q^{\bar{a}} \cap |w|^k, w[\bar{i}] = \bar{a}\} = \{w \in A^* : c_{\bar{a}}(w) \cap Q^{\bar{a}} \neq \emptyset\} = L_{\diamond_{Q^{\bar{a}}}}^{\bar{a}},$$

and thus  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$  is the Boolean algebra generated by the languages  $\bigcup_{\bar{a} \in A^k} L_{\diamond_{Q^{\bar{a}}}}^{\bar{a}}$ , where  $(Q^{\bar{a}})_{\bar{a} \in A^k}$  ranges over  $\mathcal{P}(\mathbb{N}^k)^{A^k}$ . We conclude by proving that this Boolean algebra actually coincides with  $\mathcal{B}_k$ .

**Proposition 2.5.** *For any  $k \geq 1$ , the Boolean algebra  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$  is equal to  $\mathcal{B}_k$ .*

*Proof.* The Boolean algebra  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$  is generated by the languages  $\bigcup_{\bar{a} \in A^k} L_{\diamond_{Q^{\bar{a}}}}^{\bar{a}}$ , where  $(Q^{\bar{a}})_{\bar{a} \in A^k}$  ranges over  $\mathcal{P}(\mathbb{N}^k)^{A^k}$ , while  $\mathcal{B}_k$  is generated by the languages  $L_{\diamond_Q}^{\bar{a}}$ , where  $\bar{a}$

ranges over  $A^k$  and  $Q$  ranges over subsets of  $\mathbb{N}^k$ . On the one hand, it is clear that  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$  is generated by Boolean combinations of languages of the form  $L_{\diamond_{\bar{Q}}^{\bar{a}}}$ . On the other hand, fix  $Q \subseteq \mathbb{N}^k$  and a  $k$ -tuple of letters  $\bar{a} \in A^k$ . We define the family of subsets  $(Q^{\bar{b}})_{\bar{b} \in A^k}$ , where  $Q^{\bar{a}} := Q$  and, for every  $\bar{b} \neq \bar{a}$ ,  $Q^{\bar{b}} := \emptyset$ . The set  $L_{\diamond_{\bar{Q}}^{\bar{a}}}$  is clearly the set defined by the formula  $\psi((Q^{\bar{b}})_{\bar{b} \in A^k})$  and we conclude that  $\mathcal{B}_k$  is isomorphic to  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ .  $\square$

We define  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_k^u]$ , the Boolean algebra generated by the languages  $\psi(Q)$ , where  $Q$  ranges over  $P(\mathbb{N}^k)^{A^k}$  and by the languages  $L_P$ , introduced in Definition 1.44, where  $P$  ranges over subsets of  $\mathbb{N}$ .

**Corollary 2.6.** *The Boolean algebra  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_k^u]$  is equal to  $\mathcal{B}_{0,k}$ , defined as the Boolean algebra generated by the languages  $L_P$  and  $L_{\diamond_{\bar{Q}}^{\bar{a}}}$ , where  $P$  ranges over subsets of  $\mathbb{N}$ ,  $Q$  ranges over subsets of  $\mathbb{N}^k$  and  $\bar{a}$  ranges over  $A^k$ .*

*Proof.* Direct consequence of Proposition 2.5.  $\square$

Restricting our study to only nullary numerical predicates and to uniform numerical predicates, a sub-fragments of interest is

$$\mathcal{B}\Sigma_1[\mathcal{N}^u] := \bigcup_{k \in \mathbb{N}} \mathcal{B}\Sigma_1[\mathcal{N}_k^u],$$

this is why we are interested in studying  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ , for a fixed  $k \in \mathbb{N}$ .

## 2.3 The dual space via finite colourings

In this section, we provide the first elements of study of the dual space  $X_k$ , for any  $k \geq 1$ . We start by explaining how it is possible to make an analogy between elements of the dual space and finite words. Formalizing this link, and considering a different basis of  $V_k$  constructed out of the family of all finite colourings of  $\mathbb{N}^k$  leads us to a first characterisation of  $X_k$ . We conclude the section by making this characterisation even more precise, in the case  $k = 1$ .

### 2.3.1 Colourings approach

Fix  $k \geq 1$ . We proved in Proposition 2.5 that the Boolean algebra  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$  is equal to  $\mathcal{B}_k$ . We are interested in characterizing the dual space of this Boolean algebra. In Section 2.1, Proposition 2.3, we proved that this dual space, that we denoted by  $X_k$ , is homeomorphic to the image of the continuous map  $c^k : A^* \rightarrow V_k$ , where  $V_k = \mathcal{V}(\beta(\mathbb{N}^k))^{A^k}$ . The points of  $V_k$  have a behaviour that is, in a way, similar to finite words. Let us explain the idea of this analogy in the case  $k = 1$ . A way to encode a finite word is by following the insight of logic on words: it is equivalent to consider a finite word and a family of finite disjoint subsets of  $\mathbb{N}$ , possibly empty for some of them, which

cover the initial segment  $\{0, \dots, n-1\}$ , for some integer  $n \geq 1$ . Labelling each of these subsets with a letter of the alphabet, this amounts to grouping together the positions of the word which correspond to the same letter. This is the definition of  $c_a(w)$ , where  $a \in A$  and  $w \in A^*$ . Note that, since  $V_1$  is the image of  $c^1$ , the family  $c^1(w) = (\widehat{c_a(w)})_{a \in A}$  is a point of  $V_1$ . Now, fix a point  $\vec{C} = (C_a)_{a \in A} \in V_1$ . If we view an ultrafilter  $\alpha \in \beta(\mathbb{N})$  as a generalized position, then just like in the previous situation, we could say that  $\vec{C}$  has the letter  $a \in A$  at the generalized position  $\alpha$  if, and only if,  $\alpha \in C_a$ . This reasoning applies for any  $k \geq 1$ , and for this reason, we refer to points of  $\vec{C} \in V_k$  as *generalized words*.

In the particular case where a point  $\vec{C} \in V_k$  is in the subspace  $X_k$ , we can make this analogy with words even more precise. The following characterisation of  $X_k$  is a direct consequence of the universal property of Čech-Stone compactification, and relates  $\vec{C}$  to the existence of a certain ultrafilter in  $\beta(A^*)$ . In particular, if the corresponding ultrafilter is trivial, i.e. there exists a finite word  $w \in A^*$  such that  $\gamma = \uparrow\{w\}$ , then for every  $\bar{a} \in A^k$ ,  $C_{\bar{a}}$  corresponds to the clopen associated to the set of  $k$ -tuples of positions  $c_{\bar{a}}(w)$ .

**Proposition 2.7.** *Let  $k \geq 1$  and let  $\vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k}$  be an arbitrary element of  $V_k$ . Then  $\vec{C}$  is in  $X_k$  if, and only if, there exists an ultrafilter  $\gamma \in \beta(A^*)$  such that, for every  $\bar{a} \in A^k$ , and for every  $Q \subseteq \mathbb{N}^k$ ,*

$$C_{\bar{a}} \in \square^{\bar{a}}(\hat{Q}) \text{ if, and only if, } L_{\square^{\bar{a}}_Q} \in \gamma;$$

or, equivalently,

$$C_{\bar{a}} \in \diamond^{\bar{a}}(\hat{Q}) \text{ if, and only if, } L_{\diamond^{\bar{a}}_Q} \in \gamma.$$

*Proof.* By Proposition 2.3, a generalized word  $\vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k}$  is in  $X_k$  if, and only if, it is in  $Im(c^k)$ , that is if there exists an ultrafilter  $\gamma \in \beta(A^*)$  such that  $\vec{C} = c^k(\gamma)$ . Now, by definition,  $\vec{C} = c^k(\gamma)$  if, and only if, for every  $\bar{a} \in A^k$ ,  $C_{\bar{a}} = c_{\bar{a}}(\gamma)$ . We make two more observations. On the one hand, by Remark 1.22 we know that

$$c_{\bar{a}}(\gamma) = \bigcap_{\substack{Q \subseteq \mathbb{N}^k \\ L_{\square^{\bar{a}}_Q} \in \gamma}} \hat{Q}. \quad (2.1)$$

On the other hand,

$$C_{\bar{a}} = \bigcap_{\substack{Q \subseteq \mathbb{N}^k \\ C_{\bar{a}} \subseteq \hat{Q}}} \hat{Q} = \bigcap_{\substack{Q \subseteq \mathbb{N}^k \\ C_{\bar{a}} \in \square^{\bar{a}}(\hat{Q})}} \hat{Q}. \quad (2.2)$$

We now prove that, for every  $\bar{a} \in A^k$ ,  $C_{\bar{a}} = c_{\bar{a}}(\gamma)$  if, and only if,

$$\text{for every } Q \subseteq \mathbb{N}^k, (C_{\bar{a}} \in \square^{\bar{a}}(\hat{Q}) \text{ if, and only if, } L_{\square^{\bar{a}}_Q} \in \gamma),$$

which will allow us to conclude. Fix  $\bar{a} \in A^k$ . For the right-to-left implication, let us

assume that, for every  $Q \subseteq \mathbb{N}^k$ ,  $(C_{\bar{a}} \in \square^{\bar{a}}(\widehat{Q})$  if, and only if,  $L_{\square^{\bar{a}}_Q} \in \gamma$ ). In that case, it is clear that the right hand sides of Equation 2.1 and Equation 2.2 are equal, and we obtain that  $C_{\bar{a}} = c_{\bar{a}}(\gamma)$ . For the left-to-right implication, assume that  $C_{\bar{a}} = c_{\bar{a}}(\gamma)$ .

- First, for any  $Q \subseteq \mathbb{N}^k$ , if  $C_{\bar{a}} \in \square^{\bar{a}}(\widehat{Q})$ , this means that  $C_{\bar{a}}$  is contained in  $\widehat{Q}$ . Now, by Equation 2.1, the fact that  $C_{\bar{a}} = c_{\bar{a}}(\gamma)$ , and compactness of  $V_k$ , there exists a finite collection of subsets of  $\mathbb{N}^k$ ,  $Q_1, \dots, Q_n$ , for some  $n \geq 1$ , such that  $\bigcap_{i=1}^n \widehat{Q}_i \subseteq \widehat{Q}$  and, for all  $i \in \{1, \dots, n\}$ ,  $L_{\square^{\bar{a}}_{Q_i}} \in \gamma$ . Now, since  $\gamma$  is a filter, and since the map sending any subset  $P$  of  $\mathbb{N}^k$  to the language  $L_{\square^{\bar{a}}_P}$  preserves meets (see Proof of Proposition 2.3), we have that  $L_{\square^{\bar{a}}_{Q'}} \in \gamma$ , where  $Q' := \bigcap_{i=1}^n Q_i$ . Also, since  $\widehat{Q'} \subseteq \widehat{Q}$ , and since the map that sends a subset of  $\mathbb{N}^k$  to its hat is an order-embedding, we have that  $Q' \subseteq Q$ . Thus  $L_{\square^{\bar{a}}_Q}$  is also in  $\gamma$ .
- Second, for any  $Q \subseteq \mathbb{N}^k$ , if  $L_{\square^{\bar{a}}_Q}$  is in  $\gamma$ , then  $c_{\bar{a}}(\gamma)$  is contained in  $\widehat{Q}$ . Since  $C_{\bar{a}} = c_{\bar{a}}(\gamma)$ , then by Equation 2.2, and, again, by compactness of  $V_k$ , there exists a finite family of subsets of  $\mathbb{N}^k$ ,  $Q_1, \dots, Q_n$ , for some  $n \geq 1$ , such that  $\bigcap_{i=1}^n Q_i \subseteq Q$  and, for every  $i \in \{1, \dots, n\}$ ,  $C_{\bar{a}} \subseteq \widehat{Q}_i$ . It follows that  $C_{\bar{a}}$  is contained in  $\widehat{Q}$ , which means that  $C_{\bar{a}}$  is an element of  $\square^{\bar{a}}\widehat{Q}$ , as required.

Finally,

$$\diamond^{\bar{a}}\widehat{Q} = (\square^{\bar{a}}\widehat{Q}^c)^c \text{ and } L_{\diamond^{\bar{a}}_Q} = (L_{\square^{\bar{a}}_{Q^c}})^c$$

thus we conclude that  $\vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k}$  is in  $X_k$  if, and only if, there exists  $\gamma \in \beta(A^*)$  such that, for every  $\bar{a} \in A^k$  and every  $Q \subseteq \mathbb{N}^k$ ,

$$C_{\bar{a}} \in \diamond^{\bar{a}}(\widehat{Q}) \text{ if, and only if, } L_{\diamond^{\bar{a}}_Q} \in \gamma.$$

□

This motivates the following terminology: we refer to points  $\vec{C} \in X_k$  as *pseudofinite words*. This setting allows us to compute some elementary instances of pseudofinite words which are not finite.

**Example 2.8.** In the case where  $k = 1$  and  $|A| = \{a, b\}$ , let us consider the couple of closed subsets of  $\beta(\mathbb{N})$

$$\vec{C} = (C_a, C_b) := (\beta(\mathbb{N}), *\mathbb{N}).$$

We prove that this is a pseudofinite word by giving a description of an ultrafilter  $\gamma$  satisfying the condition introduced in Proposition 2.7. Intuitively, the ultrafilter which gives this pseudofinite word will have the letter  $a$  at every standard position however, it will also have the letter  $b$  at a non-standard position.

An ultrafilter  $\gamma \in \beta(A^*)$  satisfying the condition introduced in Proposition 2.7 has to be such that:

1.  $\{P \subseteq \mathbb{N} : L_{\square^a_P} \in \gamma\} = \{\mathbb{N}\};$



$$2. \{P \subseteq \mathbb{N} : L_{\square_P^b} \in \gamma\} = \text{Cof}(\mathbb{N}).$$

that is

1.  $L_{\square_{\mathbb{N}}^a} \in \gamma$  and, for every  $S \subsetneq \mathbb{N}$ ,  $L_{\square_S^a} \notin \gamma$ .
2. For every cofinite subset  $S$  of  $\mathbb{N}$ ,  $L_{\square_S^b} \in \gamma$  and, for every non-cofinite subset  $S$  of  $\mathbb{N}$ ,  $L_{\square_S^b} \notin \gamma$ .

Let us reformulate these conditions.

1.  $L_{\square_{\mathbb{N}}^a}$  is equal to  $A^*$ , thus the first condition always holds. We prove that the second condition is equivalent to saying that, for every  $n \in \mathbb{N}$ ,  $L_{\diamond_{\{n\}}^a} \in \gamma$ . On the one hand, assume that, for every  $S \subsetneq \mathbb{N}$ ,  $L_{\square_S^a} \notin \gamma$ . Then, in particular, for any  $n \in \mathbb{N}$ ,  $L_{\square_{\mathbb{N} \setminus \{n\}}^a} \notin \gamma$ , which is equivalent to  $L_{\diamond_{\{n\}}^a} \in \gamma$ . On the other hand, assume that, for every  $n \in \mathbb{N}$ ,  $L_{\diamond_{\{n\}}^a} \in \gamma$ . For any subset  $S$  strictly contained in  $\mathbb{N}$ , pick  $n \in S^c$ . We have that  $L_{\diamond_{\{n\}}^a} \subseteq L_{\diamond_{S^c}^a} = (L_{\square_S^a})^c$ . By upset, this last language is in  $\gamma$ , and thus we conclude that  $L_{\square_S^a} \notin \gamma$ .
2. We prove, in a similar fashion, that saying that, for every cofinite subset  $S$  of  $\mathbb{N}$ ,  $L_{\square_S^b} \in \gamma$  is equivalent to saying that, for every  $n \in \mathbb{N}$ ,  $(L_{\diamond_{\{n\}}^b})^c \in \gamma$ . We also have that saying that, for every non-cofinite subset  $S$  of  $\mathbb{N}$ ,  $L_{\square_S^b} \notin \gamma$  is equivalent to saying that, for any infinite subset  $S$  of  $\mathbb{N}$ ,  $L_{\diamond_S^b} \in \gamma$ .

We conclude that the condition  $\gamma \in \beta(A^*)$  has to satisfy can be rephrased as follows.

1. For every  $n \in \mathbb{N}$ ,  $L_{\diamond_{\{n\}}^a} \in \gamma$
2. For every  $n \in \mathbb{N}$ ,  $(L_{\diamond_{\{n\}}^b})^c \in \gamma$  and for every infinite subset  $S \subseteq \mathbb{N}$ ,  $L_{\diamond_S^b} \in \gamma$ .

Put in more simple terms, the idea is the following: the ultrafilter  $\gamma$  exists because for every length  $l$  and every infinite subset  $S$ , there exists a finite word satisfying the condition "every letter up until  $l$  is  $a$ , and there exists an occurrence of  $b$  in the subset  $S$ ."

Now, for any  $n_1, \dots, n_{l_1}, m_1, \dots, m_{l_2} \in \mathbb{N}$ , where  $l_1, l_2 \geq 1$ , and for any finite family of infinite subsets  $S_1, \dots, S_l$ , the language

$$\bigcap_{i=1}^{l_1} L_{\diamond_{\{n_i\}}^a} \cap \bigcap_{i=1}^{l_2} (L_{\diamond_{\{m_i\}}^b})^c \cap \bigcap_{i=1}^l L_{\diamond_{S_i}^b}$$

is non-empty, thus

$$\uparrow \{L_{\diamond_{\{n\}}^a}, (L_{\diamond_{\{m\}}^b})^c, L_{\diamond_S^b} : n, m \in \mathbb{N}, S \subseteq \mathbb{N} \text{ and } S \text{ infinite}\}$$

is a filter and by Stone's theorem it can be extended into an ultrafilter  $\gamma$  which satisfies, by construction, the condition introduced in Proposition 2.7.

Another, and potentially more practical, characterisation of  $X_k$ , for any  $k \geq 1$ , can be made by using the terminology of finite colouring. We could summarize this characterisation by saying that a generalized word  $\vec{C}$  is pseudofinite if, and only if, for every finite colouring, it is possible to construct an actual finite word  $w \in A^*$  which is equivalent to  $\vec{C}$  when we look at it from the perspective of this finite colouring.

**Definition 2.9.** A *finite colouring* of  $\mathbb{N}^k$  is a map  $q : \mathbb{N}^k \rightarrow I$ , where  $I$  is a finite set, or, equivalently, a finite family of pairwise disjoint subsets of  $\mathbb{N}^k$ ,  $\mathcal{Q} = (Q_i)_{i \in I}$ , such that  $\bigcup_{i \in I} Q_i = \mathbb{N}^k$ . A subset  $Q \subseteq \mathbb{N}^k$  is saturated with respect to a colouring  $\mathcal{Q}$  if there exists a finite subset  $J \subseteq I$  such that  $Q = \bigcup_{j \in J} Q_j$ .

First, let us define the notion of content of a word on a subset of  $\mathbb{N}^k$ , that is the set of  $k$ -tuples of letters of  $w$  which occur on the given subset. This generalizes the notation  $w[\bar{i}]$ , where  $\bar{i} \in \mathbb{N}^k$ , introduced in Definition 1.44.

**Definition 2.10.** For any finite word  $w \in A^*$  and any subset  $Q \subseteq \mathbb{N}^k$ , we define the *content of  $w$  on  $Q$*  as

$$\langle w, Q \rangle := \{\bar{a} \in A^k : c_{\bar{a}}(w) \cap Q \neq \emptyset\}.$$

More generally, for any finite colouring  $q : \mathbb{N}^k \rightarrow I$  of  $\mathbb{N}^k$ , we define the (color) profile of a finite word  $w$  as

$$\langle w, q \rangle := (\langle w, q^{-1}(i) \rangle)_{i \in I} \in \mathcal{P}(A^k)^I,$$

that is, the family of contents of  $w$  associated to each colour.

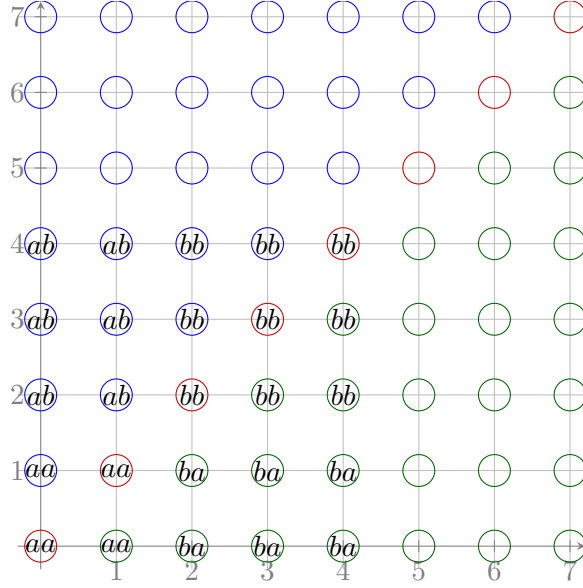
**Example 2.11.** Fix  $A = \{a, b\}$ . In the case  $k = 2$ , let us consider  $q$ , the three colours colouring  $(\Delta^<, \Delta, \Delta^>)$  of  $\mathbb{N}^2$ , where

$$\Delta^< := \{(n, m) \in \mathbb{N}^2 : n < m\},$$

$$\Delta^> := \{(n, m) \in \mathbb{N}^2 : n > m\},$$

$$\Delta := \{(n, n) \in \mathbb{N}^2 : n \in \mathbb{N}\}$$

and the finite word  $w = aabbb$ .



The profile of  $w$  for the colouring  $q$  is

$$\langle w, q \rangle = (\langle w, \Delta^< \rangle, \langle w, \Delta \rangle, \langle w, \Delta^> \rangle) = (\{aa, bb, ab\}, \{aa, bb\}, \{aa, bb, ba\}).$$

In our framework, a natural idea is to extend this notion of (colour) profile to generalized words.

**Definition 2.12.** For any subset  $Q$  of  $\mathbb{N}^k$ , and any generalized word  $\vec{C} = (C_a)_{\bar{a} \in A^k} \in V_k$ , we define the content of  $\vec{C}$  on  $Q$  as

$$\langle \vec{C}, Q \rangle := \{\bar{a} \in A^k : C_{\bar{a}} \cap \hat{Q} \neq \emptyset\}.$$

The colour profile of a generalized word  $\vec{C} = (C_a)_{\bar{a} \in A^k} \in V_k$  on a finite colouring  $q : \mathbb{N}^k \rightarrow I$  is the map

$$\langle \vec{C}, q \rangle : I \rightarrow \mathcal{P}(A^k)$$

which sends any  $i \in I$  to

$$\langle \vec{C}, q \rangle(i) := \langle \vec{C}, q^{-1}(i) \rangle.$$

In particular, if the point of  $V_k$  we consider is of the form  $(\widehat{c_{\bar{a}}(w)})_{\bar{a} \in A^k}$  for some finite word  $w \in A^*$ , then for any finite colouring  $q : \mathbb{N}^k \rightarrow I$ , we have

$$\langle (\widehat{c_{\bar{a}}(w)})_{\bar{a} \in A^k}, q \rangle = \langle w, q \rangle,$$

which shows that the profile of a generalized word can be seen as an extension of the notion of profile of a finite word. Also, notice that for any  $\vec{C} \in V_k$ , the map  $\langle \vec{C}, \cdot \rangle : \mathcal{P}(\mathbb{N}^k) \rightarrow \mathcal{P}(A^k)$  which sends a subset  $Q$  to  $\langle \vec{C}, Q \rangle$  is finitely additive: for any finite family of subsets  $Q_1, \dots, Q_n \subseteq \mathbb{N}^k$ ,

$$\begin{aligned}
\bigcup_{i=1}^n \langle \vec{C}, Q_i \rangle &= \bigcup_{i=1}^n \{\bar{a} \in A^k : C_{\bar{a}} \cap \widehat{Q}_i \neq \emptyset\} \\
&= \{\bar{a} \in A^k : C_{\bar{a}} \cap \bigcup_{i=1}^n \widehat{Q}_i \neq \emptyset\} \\
&= \{\bar{a} \in A^k : C_{\bar{a}} \cap \widehat{\bigcup_{i=1}^n Q_i} \neq \emptyset\} \\
&= \langle \vec{C}, \bigcup_{i=1}^n Q_i \rangle.
\end{aligned}$$

In particular, the map  $\langle w, \cdot \rangle : \mathcal{P}(\mathbb{N}^k) \rightarrow \mathcal{P}(A^k)$  which sends a subset  $Q$  to  $\langle w, Q \rangle$  is finitely additive.

We are now going to use the family of all finite colourings in order to provide a different basis for the space  $V_k$ .

**Lemma 2.13.** *For any finite colouring  $q : \mathbb{N}^k \rightarrow I$  of  $\mathbb{N}^k$ , we consider the map*

$$\langle \cdot, q \rangle : V_k \rightarrow \mathcal{P}(A^k)^I$$

*which sends any  $\vec{C} \in V_k$  to  $\langle \vec{C}, q \rangle$ . We also consider the family of all preimages, for all of these maps*

$$\mathcal{C} := \{\langle \cdot, q \rangle^{-1}(\bar{B}) : q : \mathbb{N}^k \rightarrow I, \text{ where } I \text{ is a finite set, and } \bar{B} \in \mathcal{P}(A^k)^I\}.$$

*The following statements hold.*

(1): *The inverse image of any point  $\bar{B} = (B_i)_{i \in I} \in \mathcal{P}(A^k)^I$  under  $\langle \cdot, q \rangle$  is clopen in  $V_k$ . In particular, for any finite colouring  $q : \mathbb{N}^k \rightarrow I$  of  $\mathbb{N}^k$ , the map  $\langle \cdot, q \rangle$  is continuous when  $\mathcal{P}(A^k)^I$  is equipped with the discrete topology.*

(2): *Any intersection of two elements in  $\mathcal{C}$  can be written as a finite union of elements in  $\mathcal{C}$ . In particular,  $\mathcal{C}$  is a basis for the topology on  $V_k$ .*

*Proof.* (1): Fix  $q : \mathbb{N}^k \rightarrow I$  a finite colouring of  $\mathbb{N}^k$ , and a family of subsets  $\bar{B} = (B_i)_{i \in I} \in \mathcal{P}(A^k)^I$ . Recall that, for any  $Q \subseteq \mathbb{N}^k$ , the subsets of the form

$$\diamond Q = \{C \in \mathcal{V}(\beta(\mathbb{N}^k)) : C \cap \widehat{Q} \neq \emptyset\}$$

are clopen in  $\mathcal{V}(\beta(\mathbb{N}^k))$ . We can express  $\langle \cdot, q \rangle^{-1}(\bar{B})$  as a finite Boolean combination of

these clopen subsets:

$$\begin{aligned}
\langle \cdot, q \rangle^{-1}(\bar{B}) &= \{ \vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k} \in V_k : \langle \vec{C}, q \rangle = \bar{B} \} \\
&= \{ \vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k} \in V_k : \forall i \in I, \langle \vec{C}, q \rangle(i) = B_i \} \\
&= \{ \vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k} \in V_k : \forall i \in I, \forall \bar{a} \in A^k, (C_{\bar{a}} \cap \widehat{q^{-1}(i)} \neq \emptyset \iff \bar{a} \in B_i) \} \\
&= \bigcap_{i \in I} \left( \bigcap_{\bar{a} \in B_i} p_{\bar{a}}^{-1}(\widehat{q^{-1}(i)}) \cap \bigcap_{\bar{a} \notin B_i} p_{\bar{a}}^{-1}(\widehat{q^{-1}(i)})^c \right),
\end{aligned}$$

where, for every  $\bar{a} \in A^k$ ,  $p_{\bar{a}} : V_k \rightarrow \mathcal{V}(\beta(\mathbb{N}^k))$  sends any  $\vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k}$  to  $C_{\bar{a}}$ . Since  $\mathcal{P}(A^k)^I$  is equipped with the discrete topology, this proves that the map  $\langle \cdot, q \rangle$  is continuous.

(2): Fix two finite colourings  $\mathcal{Q} = (Q_1, \dots, Q_l)$  and  $\mathcal{Q}' = (Q'_1, \dots, Q'_n)$  of  $\mathbb{N}^k$ , where  $l, n \geq 1$ . Fix  $\bar{B} = (B_1, \dots, B_l) \in \mathcal{P}(A^k)^l$  and  $\bar{B}' = (B'_1, \dots, B'_n) \in \mathcal{P}(A^k)^n$ . First, we define a finite colouring of  $\mathbb{N}^k$  which refines both  $\mathcal{Q}$  and  $\mathcal{Q}'$ :  $\mathcal{R} = (R_{i,j})_{\substack{1 \leq i \leq l \\ 1 \leq j \leq n}}$ , the finite colouring of  $\mathbb{N}^k$ , such that, for every  $(i, j) \in \{1, \dots, l\} \times \{1, \dots, n\}$ ,

$$R_{i,j} := Q_i \cap Q'_j.$$

Finally, we define  $\mathcal{D}_{\bar{B}, \bar{B}'} \subseteq \mathcal{P}(A^k)^{l \cdot n}$  as follows:  $\bar{D} = (D_{i,j})_{\substack{1 \leq i \leq l \\ 1 \leq j \leq n}}$  is in  $\mathcal{D}_{\bar{B}, \bar{B}'}$  if, and only if, for every  $i \in \{1, \dots, l\}$ ,

$$\bigcup \{ D_{u,v} : (u, v) \in \{1, \dots, l\} \times \{1, \dots, n\} \text{ and } R_{u,v} \subseteq Q_i \} = B_i$$

and for every  $j \in \{1, \dots, n\}$ ,

$$\bigcup \{ D_{u,v} : (u, v) \in \{1, \dots, l\} \times \{1, \dots, n\} \text{ and } R_{u,v} \subseteq Q'_j \} = B'_j.$$

We now prove that

$$\langle \cdot, \mathcal{Q} \rangle^{-1}(\bar{B}) \cap \langle \cdot, \mathcal{Q}' \rangle^{-1}(\bar{B}') = \bigcup_{\bar{D} \in \mathcal{D}_{\bar{B}, \bar{B}'}} \langle \cdot, \mathcal{R} \rangle^{-1}(\bar{D}).$$

For the left-to-right inclusion, fix  $\vec{C} \in \langle \cdot, \mathcal{Q} \rangle^{-1}(\bar{B}) \cap \langle \cdot, \mathcal{Q}' \rangle^{-1}(\bar{B}')$ . We define  $\bar{D}_{\vec{C}} \in \mathcal{P}(A^k)^{l \cdot n}$  as follows: set, for any  $(u, v) \in \{1, \dots, l\} \times \{1, \dots, n\}$ ,

$$(\bar{D}_{\vec{C}})_{u,v} := \langle \vec{C}, R_{u,v} \rangle.$$

By construction, it is clear that  $\vec{C}$  belongs to  $\langle \cdot, \mathcal{R} \rangle^{-1}(\bar{D}_{\vec{C}})$ . All we have left to prove in order to conclude is that  $\bar{D}_{\vec{C}}$  is in  $\mathcal{D}_{\bar{B}, \bar{B}'}$ . By definition of  $\bar{D}_{\vec{C}}$  we have that, for any  $i \in \{1, \dots, l\}$ ,

$$\bigcup_{\substack{(u,v) \in \{1, \dots, l\} \times \{1, \dots, n\} \\ R_{u,v} \subseteq Q_i}} (D_{\vec{C}})_{u,v} = \bigcup_{\substack{(u,v) \in \{1, \dots, l\} \times \{1, \dots, n\} \\ R_{u,v} \subseteq Q_i}} \langle \vec{C}, R_{u,v} \rangle.$$

Now, since the map  $\langle \cdot, \mathcal{R} \rangle : \mathcal{P}(\mathbb{N}^k) \rightarrow \mathcal{P}(A^k)^{l,n}$  is finitely additive, this is also equal to

$$\langle \vec{C}, \bigcup_{\substack{(u,v) \in \{1, \dots, l\} \times \{1, \dots, n\} \\ R_{u,v} \subseteq Q_i}} R_{u,v} \rangle,$$

that is,  $\langle \vec{C}, Q_i \rangle$  and since  $\vec{C}$  is in  $\langle \cdot, \mathcal{Q} \rangle^{-1}(\bar{B})$  we finally obtain that

$$\bigcup_{\substack{(u,v) \in \{1, \dots, l\} \times \{1, \dots, n\} \\ R_{u,v} \subseteq Q_i}} (D_{\vec{C}})_{u,v} = B_i.$$

We prove in the exact same way that, for any  $j \in \{1, \dots, n\}$ ,

$$\bigcup_{\substack{(u,v) \in \{1, \dots, l\} \times \{1, \dots, n\} \\ R_{u,v} \subseteq Q_i}} (D_{\vec{C}})_{u,v} = B'_j,$$

which ends to prove that  $\bar{D}_{\vec{C}} \in \mathcal{D}_{\bar{B}, \bar{B}'}$  and allows us to conclude.

For the left-to-right inclusion, fix  $\bar{D}_{\vec{C}} \in \mathcal{D}_{\bar{B}, \bar{B}'}$ . For any  $\vec{C} \in \langle \cdot, \mathcal{R} \rangle^{-1}(\bar{D})$ , we have that, for every  $i \in \{1, \dots, l\}$ ,

$$\langle \vec{C}, Q_i \rangle = \langle \vec{C}, \bigcup_{\substack{(u,v) \in \{1, \dots, l\} \times \{1, \dots, n\} \\ R_{u,v} \subseteq Q_i}} R_{u,v} \rangle.$$

Now, since the map  $\langle \cdot, \mathcal{R} \rangle : \mathcal{P}(\mathbb{N}^k) \rightarrow \mathcal{P}(A^k)^{l,n}$  is finitely additive, this is also equal to

$$\bigcup_{\substack{(u,v) \in \{1, \dots, l\} \times \{1, \dots, n\} \\ R_{u,v} \subseteq Q_i}} \langle \vec{C}, R_{u,v} \rangle$$

and since  $\bar{D} \in \mathcal{D}_{\bar{B}, \bar{B}'}$ , this is equal to

$$\bigcup_{\substack{(u,v) \in \{1, \dots, l\} \times \{1, \dots, n\} \\ R_{u,v} \subseteq Q_i}} D_{u,v}$$

which allows us to conclude that  $\langle \vec{C}, Q_i \rangle = B_i$ . The exact same reasoning can be conducted to prove that  $\langle \cdot, \mathcal{R} \rangle^{-1}(\bar{D}) \subseteq \langle \cdot, \mathcal{Q}' \rangle^{-1}(\bar{B}')$ , which allows us to conclude.

Finally, we prove that  $\mathcal{C}$  is a basis for the topology on  $V_k$ . Since  $V_k$  has been defined as a product topology, a subbasis  $\mathcal{S}$  for  $V_k$  is given, by definition, by inverse images of projections on Vietoris of  $\beta(\mathbb{N}^k)$ . We already proved in the first part of the Lemma that, for any finite colouring  $q : \mathbb{N}^k \rightarrow I$  of  $\mathbb{N}^k$ , the map  $\langle \cdot, q \rangle$  is continuous with respect to the topology on  $V_k$ . This means that every set in  $\mathcal{C}$  is open. We only have left to prove that each subbasic open in  $\mathcal{S}$  is in the topology generated by  $\mathcal{C}$ . For this, let  $\bar{a}$  be an arbitrary  $k$ -tuple of letters of  $A$  and  $Q$  be an arbitrary subset of  $\mathbb{N}^k$  giving a generic clopen  $\hat{Q}$  of the  $\bar{a}$ -th copy of  $\beta(\mathbb{N}^k)$ . We denote by  $U$  the inverse image of this clopen under the projection onto the  $\bar{a}$ -th coordinate. A point  $\vec{C} \in V_k$  is in  $U$  if, and only if,  $C_{\bar{a}} \in \square(\hat{Q})$ ,

that is, if, and only if,  $C_{\bar{a}} \subseteq \widehat{Q}$ . Now, let us consider the 2-colors coloring  $q : \mathbb{N}^k \rightarrow \{0, 1\}$ , where  $q^{-1}(0) := Q$  and  $q^{-1}(1) := Q^c$ . By the preceding chain of equivalences, we have that  $\vec{C} \in U$  if, and only if,  $\langle \vec{C}, q \rangle(0)$  contains the element  $\bar{a}$  and  $\langle \vec{C}, q \rangle(1)$  does not contain the element  $\bar{a}$ . This gives the finite union of sets in  $\mathcal{C}$  we wanted, which allows us to conclude.  $\square$

We use this basis for the topology on  $V_k$  in order to characterise  $X_k$ : the pseudofinite words  $\vec{C}$  in  $V_k$  are exactly the points such that, for each finite colouring  $q$ , we can construct a concrete finite word  $w_q \in A^*$  which has the same profile than  $\vec{C}$  on  $q$ .

**Proposition 2.14.** *A generalized word  $\vec{C} \in V_k$  is pseudofinite if, and only if, for every finite colouring  $q$  of  $\mathbb{N}^k$ , there exists a finite word  $w_q \in A^*$  such that the profiles of  $\vec{C}$  and  $w_q$  on  $q$  coincide. In particular, for any subset  $Q$  of  $\mathbb{N}^k$  which is saturated with respect to  $q$ , we have, for every  $\bar{a} \in A^k$ , that*

$$C_{\bar{a}} \cap \widehat{Q} \neq \emptyset \text{ if, and only if, } c_{\bar{a}}(w_q) \cap Q \neq \emptyset.$$

*Proof.* A generalized word  $\vec{C} \in V_k$  is pseudofinite if, and only if, it is in  $X_k$ . Recall that a  $X_k$  is the closure of the image of  $A^*$  under the map  $c^k : A^* \rightarrow V_k$ , which sends a finite word  $w \in A^*$  to  $(\widehat{c_{\bar{a}}(w)})_{\bar{a} \in A^k}$ . We proved in Lemma 2.13 (2), that the family

$$\mathcal{C} = \{ \langle \cdot, q \rangle^{-1}(\bar{B}) : \bar{B} \in \mathcal{P}(A^k)^I \text{ and } q : \mathbb{N}^k \rightarrow I, \text{ where } I \text{ is a finite set} \}$$

forms a basis for the topology on  $V_k$ . Therefore, the characterization of topological closure by a basis provides the following characterisation of  $X_k$ : a generalized word  $\vec{C} \in V_k$  is in  $X_k$  if, and only if, for every finite colouring  $q : \mathbb{N}^k \rightarrow I$  of  $\mathbb{N}^k$ , and every  $\bar{B} \in \mathcal{P}(A^k)^I$  such that  $\langle \vec{C}, q \rangle = \bar{B}$ , we have

$$\langle \cdot, q \rangle^{-1}(\bar{B}) \cap c^k(A^*) \neq \emptyset.$$

Note that this last condition is equivalent to saying that there exists a finite word  $w_q$  such that

$$\langle (\widehat{c_{\bar{a}}(w_q)})_{\bar{a} \in A^k}, q \rangle = \bar{B}.$$

We previously observed that the profile of  $(\widehat{c_{\bar{a}}(w_q)})_{\bar{a} \in A^k}$  on  $q$  is the profile of  $w_q$  on  $q$ . We conclude that  $\vec{C} \in V_k$  is pseudofinite if, and only if, there exists a finite word  $w_q$  such that  $\langle \vec{C}, q \rangle = \langle w_q, q \rangle$ .

The other statement is a direct consequence of the fact that, for any finite word  $w \in A^*$ , the map  $\langle w, \cdot \rangle : \mathcal{P}(\mathbb{N}^k) \rightarrow \mathcal{P}(A^k)$ , which sends a subset  $Q$  to  $\langle w, Q \rangle$ , and for any  $\vec{C} \in V_k$ , the map  $\langle \vec{C}, \cdot \rangle : \mathcal{P}(\mathbb{N}^k) \rightarrow \mathcal{P}(A^k)$  which sends a subset  $Q$  to  $\{\bar{a} \in A^k : C_{\bar{a}} \cap \widehat{Q} \neq \emptyset\}$  are finitely additive.  $\square$

*Remark 2.15.* In particular, for any pseudofinite word  $\vec{C} \in X_k$ , let us consider a finite colouring  $q : \mathbb{N}^k \rightarrow I$  such that one of the colours corresponds to a singleton, that is

$$q^{-1}(i) = \{\bar{p}\}$$

for some  $i \in I$  and  $\bar{p} = (p_1, \dots, p_k) \in \mathbb{N}^k$ . In that case, we observe that any word  $w$  satisfying the condition from Proposition 2.14 is necessarily such that

$$|w| > \max\{p_j : j \in \{1, \dots, k\}\}$$

and such that, for any  $\bar{a} \in A^k$ ,

$$\bar{p} \in \text{Cont}(C_{\bar{a}}) \text{ if, and only if, } w[\bar{p}] = \bar{a},$$

where  $\text{Cont}(C_{\bar{a}})$  has been introduced in Definition 1.31. Indeed, for any  $\bar{a} \in A^k$ ,

$$\begin{aligned} \bar{p} \in \text{Cont}(C_{\bar{a}}) &\iff \bar{p} \in C_{\bar{a}} \cap \mathbb{N}^k \\ &\iff \bar{a} \in \{\bar{b} \in A^k : C_{\bar{b}} \cap \{\bar{p}\} \neq \emptyset\} \\ &\iff \bar{a} \in \langle w, q \rangle(i) \text{ by Proposition 2.14} \\ &\iff w[\bar{p}] = \bar{a}. \end{aligned}$$

This remark will come handy in the proof of Lemma 2.17.

**Example 2.16.** In the case  $k = 1$  and  $|A| = \{a, b\}$  let us consider the family of closed subset of  $\beta(\mathbb{N})$

$$\vec{C} := (\beta(\mathbb{N}), \emptyset).$$

Intuitively, this should be a pseudofinite word, that we could see as a generalization of the profinite word  $a^\omega$ . In practice, we can apply Proposition 2.14: for any finite colouring  $q : \mathbb{N} \rightarrow I$ , where  $I$  is a finite set, we set, for every  $i \in I$ ,  $n_i := \min(q^{-1}(i))$  and

$$N := \max_{i \in I} n_i.$$

We consider the word  $w_q := a^N$ . This allows us to prove that this family of closed subsets is a pseudofinite word. On the one hand, the only letter occuring in  $w_q$  is  $a$ . By construction of  $N$ , we have that, for any  $i \in I$ ,  $\langle w, q \rangle(i) = \{a\}$ . On the other hand,  $C_a = \beta\mathbb{N}$  and  $C_b = \emptyset$ . This means that, for any  $i \in I$ ,  $b \notin \langle \vec{C}, q \rangle(i)$  and  $a \in \langle \vec{C}, q \rangle(i)$ : thus  $\langle \vec{C}, q \rangle(i) = \{a\}$ .

### 2.3.2 Explicit characterisation of $X_1$

In the case where  $k = 1$ , it is not too difficult to directly simplify Proposition 2.14 into a condition that does not require us to look at every finite colouring of  $\mathbb{N}^k$ . In order



to do so, we start by proving a necessary condition that holds for every  $\vec{C} \in X_k$ . The intuition is the following. Pseudofinite words share similarities with finite words, but at the generalized level of ultrafilters. In particular, if  $\vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k} \in X_k$ , then for any  $\bar{a} \in A^k$ , the content of  $C_{\bar{a}}$ , introduced in Definition 1.31, is a subset of  $\mathbb{N}^k$ . It should be possible to view the elements of this subset as  $k$ -tuples of positions of a concrete word, with a length that is possibly infinite. We formalize this intuition here-below.

**Lemma 2.17.** *Fix  $\vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k} \in X_k$  a pseudofinite word. Then, the following statements hold.*

- (1): *For any  $\bar{p} = (p_1, \dots, p_k) \in \mathbb{N}^k$  and any  $\bar{a} = (a_1, \dots, a_k) \in A^k$ , we have that  $\bar{p} \in \text{Cont}(C_{\bar{a}})$  if, and only if, for every  $j \in \{1, \dots, k\}$ ,  $(p_j, \dots, p_j) \in \text{Cont}(C_{a_j, \dots, a_j})$ .*
- (2): *For every  $a \in A$ , consider the subset of  $\mathbb{N}$*

$$C_a^{\mathbb{N}} := \pi_1(\text{Cont}(C_{a, \dots, a})),$$

where  $\pi_1 : \mathbb{N}^k \rightarrow \mathbb{N}$  is the canonical projections on the first coordinate. Then  $(C_a^{\mathbb{N}})_{a \in A}$  is a finite colouring of a downset of  $\mathbb{N}$ , and, for every  $\bar{a} \in A^k$ ,

$$\text{Cont}(C_{\bar{a}}) = \prod_{j=1}^k C_{a_j}^{\mathbb{N}}.$$

*Proof.* (1): Fix  $\bar{p} = (p_1, \dots, p_k) \in \mathbb{N}^k$ , and  $\bar{a} = (a_1, \dots, a_k) \in A^k$ . To keep the notations concise, we set, for every  $j \in \{1, \dots, k\}$ ,

$$\bar{p}^j := (p_j, \dots, p_j) \text{ and } \bar{a}^j := (a_j, \dots, a_j).$$

We consider the following colouring of  $\mathbb{N}^k$  into  $k + 2$  colours

$$\mathcal{Q}_{\bar{p}} := (\{\bar{p}\}, \{\bar{p}^1\}, \dots, \{\bar{p}^k\}, \mathbb{N}^k \setminus \{\bar{p}, \bar{p}^1, \dots, \bar{p}^k\}).$$

Here, we assume that  $\bar{p}$  is not in  $\{\bar{p}^j : j \in \{1, \dots, k\}\}$ , but it should be noted that our proof still works when it is the case and that  $\mathcal{Q}_{\bar{p}}$  has only  $k + 1$  colours. Since  $\vec{C}$  is in  $X_k$ , by Proposition 2.14, we can consider a finite word  $w \in A^*$  which has the same profile than  $\vec{C}$  for the colouring  $\mathcal{Q}_{\bar{p}}$ . In particular,  $|w| > \max\{p_j : j \in \{1, \dots, k\}\}$ . We now prove the desired equivalence by using Remark 2.15,

$$\begin{aligned} \bar{p} \in \text{Cont}(C_{\bar{a}}) &\iff w[\bar{p}] = \bar{a} \\ &\iff \forall j \in \{1, \dots, k\}, w_{p_j} = a_j \\ &\iff \forall j \in \{1, \dots, k\}, w[\bar{p}^j] = \bar{a}^j \\ &\iff \bar{p}^j \in \text{Cont}(C_{\bar{a}^j}). \end{aligned}$$

(2): First, we prove that for any  $a, b \in A$  distincts,  $C_a^{\mathbb{N}} \cap C_b^{\mathbb{N}}$  is empty. Let us assume that  $C_a^{\mathbb{N}} \cap C_b^{\mathbb{N}}$  is non-empty. Pick an element  $l \in C_a^{\mathbb{N}} \cap C_b^{\mathbb{N}}$ , and then pick  $\bar{p} \in \text{Cont}(C_{a, \dots, a})$

and  $\bar{m} \in \text{Cont}(C_{b,\dots,b})$  such that  $p_1 = m_1 = l$ . This is possible, since  $C_a^{\mathbb{N}}$  is the projection of  $C_{a,\dots,a}$  on its first coordinate, and the same can be said for  $C_b^{\mathbb{N}}$  and  $C_{b,\dots,b}$ . We now consider the colouring of  $\mathbb{N}^k$  into three colors

$$\mathcal{Q}_{\bar{p},\bar{m}} := (\{\bar{p}\}, \{\bar{m}\}, \mathbb{N}^k \setminus \{\bar{p}, \bar{m}\}).$$

Since  $\vec{C}$  is in  $X_k$ , by Proposition 2.14, we can consider a finite word  $w \in A^*$  which has the same profile than  $\vec{C}$  for the colouring  $\mathcal{Q}_{\bar{p},\bar{m}}$ . In particular, by Remark 2.15, since  $\bar{p} \in \text{Cont}(C_{a,\dots,a})$  and  $\bar{m} \in \text{Cont}(C_{b,\dots,b})$ , we have that

$$w[\bar{p}] = (a, \dots, a) \text{ and } w[\bar{m}] = (b, \dots, b).$$

Now, since  $p_1 = m_1$ , we have that  $a = b$ , which allows us to conclude.

Finally, we prove that  $\bigcup_{a \in A} C_a^{\mathbb{N}}$  is a downset of  $\mathbb{N}$ . Fix  $a \in A$ ,  $l \in C_a^{\mathbb{N}}$  and consider some element  $n < l$ . We prove that there exists  $b \in A$  such that  $n \in C_b^{\mathbb{N}}$ . We use the notation  $\bar{n} := (n, \dots, n) \in \mathbb{N}^k$ . Picking an element  $\bar{p} \in C_{a,\dots,a}$  with  $p_1 = l$ , we consider the colouring of  $\mathbb{N}^k$  into three colors

$$\mathcal{Q}_{\bar{p},\bar{n}} := (\{\bar{p}\}, \{\bar{n}\}, \mathbb{N}^k \setminus \{\bar{p}, \bar{n}\}).$$

Since  $\vec{C}$  is in  $X_k$ , by Proposition 2.14, we can consider a finite word  $w \in A^*$  which has the same profile than  $\vec{C}$  for the colouring  $\mathcal{Q}_{\bar{p},\bar{n}}$ . In particular, by Remark 2.15,  $|w| > n$  and we can set  $b := w_n$ . We have that  $\bar{n} \in C_{b,\dots,b}$ , and thus  $n \in C_b^{\mathbb{N}}$ , which allows us to conclude.  $\square$

This condition is actually sufficient to characterise all pseudofinite words in the case  $k = 1$ .

**Proposition 2.18** (Explicit description of  $X_1$ ). *A generalized word  $\vec{C} = (C_a)_{a \in A} \in V_1$  is pseudofinite if, and only if,  $(\text{Cont}(C_a))_{a \in A}$  is a finite colouring of a downset of  $\mathbb{N}$ .*

*Proof.* The left-to-right implication is exactly Lemma 2.17, with  $k = 1$ . For the right-to-left implication, let us consider a generalized word  $\vec{C} = (C_a)_{a \in A} \in V_1$  such that  $(\text{Cont}(C_a))_{a \in A}$  is a finite colouring of a downset of  $\mathbb{N}$ . In order to conclude, we prove that  $\vec{C}$  satisfies the condition introduced in Proposition 2.14. Fix a finite colouring  $q : \mathbb{N} \rightarrow I$  of  $\mathbb{N}$ . We are going to construct a word  $w_q$  such that the profiles of  $\vec{C}$  and  $w_q$  coincide for the colouring  $q$ . First, in the case where  $\bigcup_{a \in A} \text{Cont}(C_a)$  is finite, the word  $w_q$  such that, for every  $a \in A$ ,

$$c_a(w_q) = \text{Cont}(C_a)$$

is finite, and has, by construction, the same profile than  $\vec{C}$  on  $q$ . Now, we treat the case where  $\bigcup_{a \in A} \text{Cont}(C_a)$  is equal to  $\mathbb{N}$ : we need to make sure that the word  $w_q$  we construct is long enough. Pick  $n \in \mathbb{N}$  such that any colour occurring finitely many

times does not occur after  $n$ , that is,

$$n > \max \bigcup_{\substack{i \in I \\ q^{-1}(i) \text{ finite}}} q^{-1}(i).$$

Now, for every colour  $i \in I$  that occurs, i.e.  $q^{-1}(i)$  is non-empty, and for every  $a \in A$  such that  $\text{Cont}(C_a) \cap q^{-1}(i)$  is non-empty, pick  $m_{i,a}$  in that set, and then pick

$$m > \max\{m_{i,a} : i \in I \text{ and } a \in A \text{ such that } \text{Cont}(C_a) \cap q^{-1}(i) \neq \emptyset\}.$$

Finally, set

$$l := \max(m, n).$$

We now define  $w_q$  as the word of length  $l$  that has the letter  $a$  at the position  $p < l$  if, and only if,  $p \in \text{Cont}(C_a)$ . Finally, we check that the word  $w_q$  we constructed has the same profile than  $\vec{C}$  for the colouring  $q$ . Since  $l > n$ , it is clear that  $w_q$  contains all positions corresponding to a finite colour. For an infinite colour, we make a case distinction. If  $a \in A$  is such that  $\text{Cont}(C_a) \cap q^{-1}(i)$  is non-empty, then since  $l > m_{i,a}$ ,  $w_q$  contains a position  $m_{i,a} \in \text{Cont}(C_a) \cap q^{-1}(i)$  such that the associated letter is the letter  $a$ . Otherwise, if  $a \in A$  is such that  $\text{Cont}(C_a) \cap q^{-1}(i)$  is empty, then for any position  $p < l$  in  $w_q$ ,  $p \in \text{Cont}(C_b)$  necessarily implies that  $b \neq a$ . Therefore, the profiles of  $\vec{C}$  and  $w_q$  coincide on  $q$ , and we conclude that  $\vec{C} \in X_1$ .  $\square$

## 2.4 Projection of $X_k$ on $\mathcal{V}(\beta(\mathbb{N}^k))$

In this section, we take a first step in the direction of giving a more explicit reformulation of Proposition 2.14. This proposition involved a condition that required us to quantify over every finite colouring of  $\mathbb{N}^k$ , making it a bit abstract to use in practice. Fix a letter  $a \in A$ . We try to characterise the image of  $X_k$  under the canonical projection

$$p_{a,\dots,a} : V_k \rightarrow \mathcal{V}(\beta(\mathbb{N}^k))$$

which sends every family of closed subsets  $\vec{C} = (C_{\bar{a}})_{\bar{a} \in A^k}$  to  $C_{a,\dots,a}$ . We fix  $k = 2$ , in order to keep lighter notations. We start by explaining how the results from the previous section can be reformulated in order to characterise the image of  $p_{a,a} \circ c^2$  and provide several necessary conditions that have to be satisfied by a closed subset of  $\mathcal{V}(\beta(\mathbb{N}^2))$  in order to be in the image of  $p_{a,a} \circ c^2$ .

### 2.4.1 The Boolean algebra $\mathcal{B}_2^{a,a}$ and the interior operator $R$

The Boolean algebra  $\mathcal{B}_k$  we have been studying is generated by the languages  $L_{\square \bar{a} Q}$ , where  $\bar{a}$  ranges over  $A^k$  and  $Q$  ranges over subsets of  $\mathbb{N}^k$ . If we fix a letter  $a \in A$ , then we can consider the Boolean subalgebra

$$\mathcal{B}_k^{a,\dots,a} := \langle \{L_{\square_Q^a} : Q \subseteq \mathbb{N}^k\} \rangle_{BA},$$

We denote by  $X_k^{a,\dots,a}$  the dual space of  $\mathcal{B}_k^{a,\dots,a}$ . In particular, the canonical embedding  $\mathcal{B}_k^{a,\dots,a} \hookrightarrow \mathcal{B}_k$  induces the continuous quotient map  $X_k \twoheadrightarrow X_k^{a,\dots,a}$ . We can prove, by applying the same reasoning than in the proof of Proposition 2.3, that  $X_k^{a,\dots,a}$  is the image of the map  $c_{a,\dots,a} : \beta(A^*) \rightarrow \mathcal{V}(\beta(\mathbb{N}^k))$ , which is the unique continuous extension of the map which sends a finite word  $w \in A^*$  to  $\widehat{c_{a,\dots,a}(w)}$ . Obtaining a more explicit characterisation of the points of  $X_k^{a,\dots,a}$  should enable us to get a better understanding of  $X_k$ .

We proved in Proposition 2.14 a characterisation of  $X_k$ , for any  $k \geq 1$ , in terms of finite colourings of  $\mathbb{N}^k$ . It can be simplified to determine the points in the image of  $X_k^{a,\dots,a}$ . In order to keep compact notations, we focus on the particular case  $k = 2$ .

**Proposition 2.19.** *For any closed subset  $C \in \mathcal{V}(\beta(\mathbb{N}^2))$ , the following conditions are equivalent.*

- (1):  *$C$  is a point of  $X_2^{a,a}$ .*
- (2): *For every finite colouring  $\mathcal{Q}$  of  $\mathbb{N}^2$ , there exists a finite word  $w \in A^*$  such that, for every subset  $Q$  which is saturated with respect to the colouring  $\mathcal{Q}$ ,*

$$C \cap \widehat{Q} \neq \emptyset \text{ if, and only if, } c_{a,a}(w) \cap Q \neq \emptyset.$$

- (3): *For every finite sequence of subsets  $Q, Q_1, \dots, Q_\ell$  of  $\mathbb{N}^2$ , where  $\ell \geq 1$ , such that  $C \subseteq \widehat{Q}$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $C \cap \widehat{Q_i}$  is non-empty, there exists a finite subset  $S \subseteq \mathbb{N}$  such that  $S^2 \subseteq Q$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $S^2 \cap Q_i$  is non-empty.*

*Proof.* We prove that (1) and (3) are equivalent. Recall that  $X_2^{a,a}$  is the closure of the image of  $A^*$  under the map  $c_{a,a} : A^* \rightarrow \mathcal{V}(\beta(\mathbb{N}^2))$ , which sends a finite word  $w \in A^*$  to  $\widehat{c_{a,a}(w)}$ . In Definition , we provided a basis for the topology on Vietoris hyperspaces. In particular, the family

$$\mathcal{D}_2^{a,a} := \{ \square^{a,a}(\widehat{Q}) \cap \bigcap_{i=1}^l \diamond^{a,a}(\widehat{Q_i}) : l \in \mathbb{N}, Q, Q_1, \dots, Q_\ell \subseteq \mathbb{N}^2 \}$$

forms a basis for the topology on  $\mathcal{V}(\beta(\mathbb{N}^2))$ . By general topology, the characterisation of the points  $C \in X_2^{a,a}$  in terms of this basis is the following:  $C$  is in  $X_2^{a,a}$  if, and only if, for every  $K \in \mathcal{D}_2^{a,a}$  which contains  $C$ , there exists a finite word  $w \in A^*$  such that  $\widehat{c_{a,a}(w)}$  is also in  $K$ . This is equivalent to say that, for every finite sequence of subsets  $Q, Q_1, \dots, Q_\ell$  of  $\mathbb{N}^2$  such that  $C \subseteq \widehat{Q}$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $C \cap \widehat{Q_i}$  is non-empty, there exists a finite subset  $S \subseteq \mathbb{N}$  such that  $S^2 \subseteq Q$  and for every  $i \in \{1, \dots, \ell\}$ ,  $S^2 \cap Q_i$  is non-empty.

We prove that (1) implies (2). Assume that  $C \in \mathcal{V}(\beta(\mathbb{N}^2))$  is in  $X_2^{a,a}$ , that is, that there exists  $\bar{D} \in X_2$  such that  $D_{a,a} = C$ . Consider a finite colouring  $q : \mathbb{N}^2 \rightarrow I$ . By Proposition

2.14, there exists a finite word  $w_q \in A^*$  such that, for every subset  $Q \subseteq \mathbb{N}^2$  which is saturated with respect to  $q$ , the profiles of  $\bar{D}$  and  $w_q$  on  $Q$  coincide. Focusing our attention only on the couple of letters  $(a, a)$ , this means that  $w_q$  is such that,  $C \cap \widehat{Q}$  is non-empty if, and only if,  $c_{a,a}(w) \cap Q$  is non-empty.

Finally, we prove that (2) implies (3). Fix  $C \in \mathcal{V}(\beta(\mathbb{N}^2))$ , and consider a finite sequence of subsets  $Q, Q_1, \dots, Q_\ell$  of  $\mathbb{N}^2$  such that  $C \subseteq \widehat{Q}$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $C \cap \widehat{Q}_i$  is non-empty. Now, consider a finite colouring  $q : \mathbb{N}^2 \rightarrow I$  such that  $Q, Q_1, \dots, Q_\ell$  are all saturated with respect to  $q$ . Then, by applying (2), there exists a finite word  $w \in A^*$  such that  $C \subseteq \widehat{Q}$  if, and only if  $c_{a,a}(w) \subseteq Q$  is non-empty, and for every  $i \in \{1, \dots, \ell\}$ ,  $C \cap \widehat{Q}_i$  is non-empty if, and only if  $c_{a,a}(w) \subseteq Q_i$ . In particular, setting  $S := c_a(w)$ , we obtain a finite subset of  $\mathbb{N}$  such that  $S^2 \subseteq Q$  and for every  $i \in \{1, \dots, \ell\}$ ,  $S^2 \cap Q_i$  is non-empty.  $\square$

In the case  $k = 1$ , for any letter  $a \in A$  we simply had that  $\text{Im}(c_a) = \mathcal{V}(\beta(\mathbb{N}))$ . However, the characterisation is not as simple for  $k > 1$ , as we illustrate in the following example.

**Example 2.20.** We prove that the closed subset  $^*(\mathbb{N}^2)$  introduced in Section 1.1.2 cannot belong to  $X_2^{a,a}$ . In order to do so, we use Proposition 2.19 (2). We consider the colouring of  $\mathbb{N}^2$  with three colors  $(Q_1, Q_2, Q_3)$ , where we set

$$\begin{cases} Q_1 := \{(0, 0)\} \\ Q_2 := \{0\} \times (\mathbb{N} \setminus \{0\}) \\ Q_3 := (\mathbb{N} \setminus \{0\}) \times \mathbb{N} \end{cases}.$$

Note that, by Lemma 1.15,  $^*(\mathbb{N}^2) \cap \widehat{Q}_1$  is empty, while  $^*(\mathbb{N}^2) \cap \widehat{Q}_2$  and  $^*(\mathbb{N}^2) \cap \widehat{Q}_3$  are non-empty. Let us consider a finite word  $w \in A^*$  such that  $c_{a,a}(w) \cap Q_1$  is empty, while  $c_{a,a}(w) \cap \widehat{Q}_2$  and  $c_{a,a}(w) \cap \widehat{Q}_3$  are non-empty. The first condition imply that  $w_0$  cannot be the letter  $a$ . However, any of the two remaining conditions implies that  $w_0 = a$ : we conclude that there exists no word such that  $c_{a,a}(w) \cap Q_1$  is empty while  $c_{a,a}(w) \cap \widehat{Q}_2$  and  $c_{a,a}(w) \cap \widehat{Q}_3$  are non-empty. By applying Proposition 2.19 (2), we conclude that  $^*(\mathbb{N}^2)$  does not belong to  $X_2^{a,a}$ .

We want to refine the characterisation of  $X_2^{a,a}$  obtained in Proposition 2.19. Since  $X_2^{a,a}$  is contained in  $\mathcal{V}(\beta(\mathbb{N}^2))$ , the points of this dual space can be understood as closed subsets of  $\beta(\mathbb{N}^2)$ . In order to get started, we determine which of these points correspond to subsets which are clopen. This is relatively straight-forward if we rely on the intuition we have built on generalized words. First, if  $w$  is a finite word, then the image of this word under  $p_{a,a} \circ c^2$ ,

$$p_{a,a} \circ c^2(w) = p_{a,a}(\widehat{(c_{u,v}(w))_{(u,v) \in A^2}}) = \widehat{c_{a,a}(w)},$$

is clearly in  $X_2^{a,a}$ . We deduce that, for every finite subset  $S \subseteq \mathbb{N}$ , the clopen  $\widehat{S^2}$  belongs to  $X_2^{a,a}$ . Now, let us consider a clopen subset of  $\beta(\mathbb{N}^2)$  which is not contained in  $\mathbb{N}^2$ . To

be in  $X_2^{a,a}$ , this clopen needs to correspond, in a sense, to the set of positions occupied by the letter  $a$  in a word with infinite length.

We start by stating the following necessary condition that the content of every closed subset in  $X_2^{a,a}$  needs to satisfy. This is a simple consequence of Lemma 2.17, projected onto the component corresponding to the couple of letters  $(a, a)$ .

**Proposition 2.21.** *For any  $C \in X_2^{a,a}$ , there exists a subset  $T \subseteq \mathbb{N}$  such that  $\text{Cont}(C) = T^2$ .*

*Proof.* If a closed subset  $C$  is in  $X_2^{a,a}$ , then there exists  $\vec{C} = (C_{a,b})_{a,b \in A} \in X_2$  such that  $C = C_{a,a}$ . By applying Lemma 2.17, there exists a subset  $T \subseteq \mathbb{N}$  such that  $\text{Cont}(C) = T^2$ .  $\square$

We are now ready to characterise clopen subsets in  $X_2^{a,a}$ .

**Proposition 2.22.** *A clopen  $K \in \mathcal{V}(\beta(\mathbb{N}^2))$  is in  $X_2^{a,a}$  if, and only if, there exists a subset  $T$  of  $\mathbb{N}$  such that  $K = \widehat{T^2}$ .*

*Proof.* Fix a clopen  $K \in \mathcal{V}(\beta(\mathbb{N}^2))$ . We know that there exists  $Q \subseteq \mathbb{N}^2$  such that  $K = \widehat{Q}$ . On the one hand, assume that  $\widehat{Q}$  is in  $X_2^{a,a}$ . By applying Proposition 2.21, we know that there exists a subset  $T$  of  $\mathbb{N}$  such that  $Q = T^2$ . On the other hand, assume that there exists a subset  $T \subseteq \mathbb{N}$  such that  $Q = T^2$ . We prove, by using Proposition 2.19 (3) that  $\widehat{T^2}$  is in  $X_2^{a,a}$ . Consider a finite family of subsets  $Q_1, \dots, Q_\ell$  of  $\mathbb{N}^2$ , where  $\ell \geq 1$ , such that, for every  $i \in \{1, \dots, \ell\}$ ,  $\widehat{T^2} \cap \widehat{Q_i}$  is non-empty. Then, for every  $i \in \{1, \dots, \ell\}$ , pick an element  $(n_i, m_i)$  in  $T^2 \cap Q_i$ . Setting

$$S := \{n_i, m_i : i \in \{1, \dots, \ell\}\},$$

we have, by construction, that  $S^2 \subseteq T^2$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $S^2 \cap Q_i$  is non-empty. By Remark 1.21 and Proposition 2.19 (3), we conclude that  $\widehat{T^2}$  is in  $X_2^{a,a}$ .  $\square$

What is still to be determined is which are the points in  $X_2^{a,a}$  corresponding to closed subsets which are not clopen. At this point, we are already able to answer the question when the closed subset we consider is finite.

**Definition 2.23.** For any subset  $Q$  of  $\mathbb{N}^2$ , we define the diagonal of  $Q$  as

$$\Delta(Q) := \{(n, n) \in \mathbb{N}^2 : (n, n) \in Q\}.$$

In particular, we use the notation  $\Delta$  to refer to the diagonal of  $\mathbb{N}^2$ .

**Lemma 2.24.** *Fix a non-empty closed subset  $C \in X_2^{a,a}$ . For every  $Q \subseteq \mathbb{N}^2$ , if  $C \cap \widehat{Q}$  is non-empty, then necessarily  $C \cap \widehat{\Delta(Q)}$  is non-empty.*

*Proof.* Taking the negation of this statement, assume that there exists  $C \in X_2^{a,a}$  and  $Q \subseteq \mathbb{N}^2$  such that  $C \cap \widehat{Q}$  is non-empty and  $C \cap \widehat{\Delta(Q)}$  is empty, that is, that  $C \subseteq \widehat{\Delta(Q)^c}$ .

There cannot exist a finite subset  $S \subseteq \mathbb{N}$  such that  $S^2 \subseteq \Delta(Q)^c$  and  $S^2 \cap Q$  is non-empty. By Proposition 2.19 (3),  $C$  cannot be in  $X_2^{a,a}$ , which allows us to conclude.  $\square$

The previous lemma can already be used in order to tell whether some given closed subsets of  $\beta(\mathbb{N}^2)$  belong to  $X_2^{a,a}$ .

**Example 2.25.** Fix  $T$  a finite, non-empty, subset of  $\mathbb{N}$  and  $\alpha \in {}^*\mathbb{N}^2$ . Let us consider the closed subset of  $\beta(\mathbb{N}^2)$

$$C := \widehat{T^2} \cup \{\alpha\}.$$

We prove that  $C$  is not in  $X_2^{a,a}$ . Let us assume that  $C \in X_2^{a,a}$ . We are going to use Proposition 2.19 (3). Fix  $n \in T$ , and  $Q \in \alpha$ . Set  $Q_0 := \Delta(Q) \cup T^2$ ,  $Q_1 := \{(n, n)\}$  and  $Q_2 := \Delta(Q)$ . By construction, it is clear that  $C$  is contained in  $\widehat{Q_0}$  and that  $\widehat{Q_1}$  has a non-empty intersection with  $C$ . The fact that  $\widehat{Q_2}$  has a non-empty intersection with  $C$  comes from Lemma 2.24. Applying Proposition 2.19 (3), there exists a finite subset  $S \subseteq \mathbb{N}$  such that  $S^2 \subseteq Q_0$ , and such that  $S^2 \cap Q_1$  and  $S^2 \cap Q_2$  are non-empty. Fix  $(m, m)$  in  $S^2 \cap Q_2$ . Since  $m \notin T$ ,  $n$  and  $m$  are not equal. Now,  $(n, m)$  is in  $S^2$ , and thus in  $Q_0$ , which is not possible since  $n \in T$  and  $m \notin T$ . We conclude that  $C$  is not in  $X_2^{a,a}$ .

**Proposition 2.26.** *The finite closed subsets  $C \in X_2^{a,a}$  which are not contained in  $\mathbb{N}^2$  are exactly those of the form  $\{\alpha\}$ , where  $\alpha \in {}^*(\mathbb{N}^2)$  is a free ultrafilter such that, for every  $Q \subseteq \mathbb{N}^2$ , if  $Q \in \alpha$ , then  $\Delta(Q) \in \alpha$ .*

*Proof.* First, we check that the closed subsets of the form  $\{\alpha\}$ , where  $\alpha \in {}^*(\mathbb{N}^2)$  is a free ultrafilter as described in the proposition, are indeed in  $X_2^{a,a}$ . We do so by using the characterisation of Proposition 2.19 (3). Consider a finite family of subsets  $Q, Q_1, \dots, Q_\ell$ , where  $\ell \geq 1$ , which all belong to  $\alpha$ . By definition of  $\alpha$ ,  $\Delta(Q), \Delta(Q_1), \dots, \Delta(Q_\ell)$  all belong to  $\alpha$ , and thus so does their intersection. Pick a singleton contained in this intersection. It provides a finite set-theoretic square that is contained in each of the subsets  $Q, Q_1, \dots, Q_\ell$ , and allows us to conclude that  $\{\alpha\}$  is in  $X_2^{a,a}$ .

Now, fix  $C \in X_2^{a,a}$  which is not contained in  $\mathbb{N}^2$ , and assume that it is finite. By Proposition 2.21, we now that  $\text{Cont}(C)$  is a set-theoretic square. Since  $C$  is finite, and not contained in  $\mathbb{N}^2$ , this means that there exists a finite subset  $T \subseteq \mathbb{N}$  such that

$$C = \widehat{T^2} \cup \{\alpha_1, \dots, \alpha_\ell\},$$

for some  $\ell \geq 1$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $\alpha_i \in {}^*(\mathbb{N}^2)$ . Let us assume that  $\ell \geq 2$ . In what follows, we construct a finite family of subsets of  $\mathbb{N}^2$  which do not satisfy Proposition 2.19 (3), and thus prove that  $\ell$  is necessarily equal to 1.

- By applying Proposition 1.19, we can construct  $\ell$  pairwise disjoint subsets  $Q_1, \dots, Q_\ell$  of  $\mathbb{N}^2$  such that, for every  $i \in \{1, \dots, \ell\}$ ,  $Q_i \in \alpha_i$ . We can assume, without any loss of generality, that for every  $i \in \{1, \dots, \ell\}$ ,  $Q_i \cap T^2$  is empty. Indeed, for every  $i \in \{1, \dots, \ell\}$ ,  $(Q_i \setminus T^2, T^2, Q_i^c)$  is a finite partition of  $\mathbb{N}^2$ , thus exactly one of these

three subsets is in  $\alpha_i$ . We know that  $Q_i^c \notin \alpha$ , and that  $T^2 \notin \alpha_i$ , since  $\alpha_i$  is a free ultrafilter: thus  $Q_i \setminus T^2 \in \alpha_i$ .

- We prove that, for every  $i \in \{1, \dots, \ell\}$ ,  $\Delta(Q_i) \in \alpha_i$ . Fix  $i \in \{1, \dots, \ell\}$ . Since  $C \cap \widehat{Q_i}$  is non-empty, then by Lemma 2.24, we also have that  $C \cap \widehat{\Delta(Q_i)}$  is non-empty. By construction,  $\Delta(Q_i)$  has an empty intersection with  $T^2$ . Also, as it is contained in  $Q_i$ , it has an empty intersection with every  $Q_j$  with  $j \in \{1, \dots, \ell\}$  such that  $i \neq j$ , and it cannot be in the corresponding  $\alpha_j$ . We conclude that  $\Delta(Q_i) \in \alpha_i$ .
- We are now ready to apply Proposition 2.19 (3). We consider the subset

$$Q := T^2 \cup \bigcup_{i=1}^{\ell} \Delta(Q_i),$$

which is, by construction, such that  $C \subseteq \widehat{Q}$ , and the  $\ell$  disjoint subsets  $\Delta(Q_1), \dots, \Delta(Q_\ell)$ , which are such that the corresponding clopens all have a non-empty intersection with  $C$ . By Proposition 2.21 (3), there exists a finite subset  $S \subseteq \mathbb{N}$  such that  $S^2 \subseteq Q$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $S^2 \cap \Delta(Q_i)$  is non-empty. In particular, pick  $(n_1, n_1) \in S^2 \cap \Delta(Q_1)$  and  $(n_2, n_2) \in S^2 \cap \Delta(Q_2)$ . Since  $\Delta(Q_1)$  and  $\Delta(Q_2)$  are disjoint, we necessarily have that  $n_1 \neq n_2$ . Now,  $(n_1, n_2)$  has to belong to  $S^2$ . This is absurd, since it cannot belong to  $T^2$ , neither to  $\bigcup_{i=1}^{\ell} \Delta(Q_i)$ .

We conclude that  $\ell$  has to be equal to 1, and by Example 2.25, that  $T$  has to be empty.  $\square$

From now on, we assume that the closed subsets  $C \in \mathcal{V}(\beta(\mathbb{N}^2))$  we consider are not contained in  $\mathbb{N}^2$ , and infinite. As we know,  $C$  is defined by the intersection of all the clopen subsets  $\widehat{Q}$  of  $\beta(\mathbb{N}^2)$  such that  $C \subseteq \widehat{Q}$ . Let us resort, once again, to the insight obtained by looking at finite words. In the case where  $C = \widehat{c_{a,a}(w)}$  for a finite word  $w \in A^*$ , the subset  $c_{a,a}(w)$  is a finite set-theoretic square. In particular,  $c_{a,a}(w)$  cannot be contained in a subset  $Q$  of  $\mathbb{N}^2$  which contains no set-theoretic square. This remark justifies the introduction of the following operator on  $\mathcal{P}(\mathbb{N}^2)$ .

**Definition 2.27.** For any subset  $Q$  of  $\mathbb{N}^2$ , we define the following subset of  $Q$

$$R(Q) := \bigcup \{S^2 : S \text{ is finite and } S^2 \subseteq Q\}.$$

We denote by  $S_Q$  the subset of  $\mathcal{P}_{fin}(\mathbb{N})$  of the finite subsets  $S \subseteq \mathbb{N}$  such that  $S^2 \subseteq Q$ .

We first prove that  $R$  is an interior operator on  $\mathcal{P}(\mathbb{N}^2)$ , which will be helpful in order to understand the structure of the dual space  $X_2^{a,a}$ .

**Lemma 2.28.** *The map  $R : \mathcal{P}(\mathbb{N}^2) \rightarrow \mathcal{P}(\mathbb{N}^2)$  is an interior operator on  $\mathcal{P}(\mathbb{N}^2)$ , i.e it is idempotent, monotone, and for every  $Q \subseteq \mathbb{N}^2$ ,  $R(Q) \subseteq Q$ . Furthermore,  $R$  preserves finite intersections.*



*Proof.* The fact that, for every  $Q \subseteq \mathbb{N}^2$ ,  $R(Q)$  is included in  $Q$  is clear by definition. For the monotonicity, for any subsets  $Q \subseteq Q' \subseteq \mathbb{N}^2$ , we have by definition of  $\mathcal{S}_Q$  and  $\mathcal{S}_{Q'}$  that

$$\begin{aligned} R(Q) &= \bigcup \{S^2 : S \in \mathcal{S}_Q\} \\ &\subseteq \bigcup \{S^2 : S \in \mathcal{S}_{Q'}\} \\ &= R(Q'). \end{aligned}$$

Finally, we prove the idempotence. For every  $Q \subseteq \mathbb{N}^2$ , by monotonicity, the fact that  $R(Q) \subseteq Q$  implies that  $R(R(Q)) \subseteq R(Q)$ , which gives one of the two inclusions. For the other inclusion, we want to prove that  $R(Q) \subseteq R(R(Q))$ . Fix  $(n, m) \in R(Q)$ . Then there exists a finite subset  $S \subseteq \mathbb{N}$  such that  $n, m \in S$  and  $S^2 \subseteq Q$ . Since, by definition,  $R(Q)$  is the union of every finite set-theoretic square contained in  $Q$ , we have that  $S^2 \subseteq R(Q)$ , and therefore  $(n, m) \in R(R(Q))$ . The fact that  $R$  reserves finite intersections follows from the general fact that a system closed under infinite unions corresponds to an interior operator (see [20], Theorem 7.3).  $\square$

The interior operator  $R$  plays a major role in our study. Fix  $C \in X_2^{a,a}$ . We prove that, in order to define  $C$ , it is actually enough to look at the clopen subsets containing  $C$  for which the corresponding subset of  $\mathbb{N}^2$  is in the image of the operator  $R$ .

**Proposition 2.29.** *If  $C \in X_2^{a,a}$ , then for every subset  $Q \subseteq \mathbb{N}^2$  such that  $\widehat{Q}$  contains  $C$ ,  $\widehat{R(Q)}$  also contains  $C$ .*

*Proof.* Fix  $C \in X_2^{a,a}$ , and consider a subset  $Q$  of  $\mathbb{N}^2$  such that  $C \subseteq \widehat{Q}$ . Consider the colouring of  $\mathbb{N}^2$  into three colors

$$(R(Q), Q \setminus R(Q), Q^c).$$

By applying Proposition 2.19 (2), there exists a finite word  $w \in A^*$  such that, in particular,  $c_{a,a}(w) \subseteq Q$ . We want to prove that  $c_{a,a}(w) \subseteq R(Q)$ , in other words, we want to prove that for any  $(i_1, i_2) \in c_{a,a}(w)$ , there exists a finite subset  $S$  of  $\mathbb{N}$  such that  $(i_1, i_2) \in S^2$  and  $S^2 \subseteq Q$ . Now, fix  $(i_1, i_2) \in \mathbb{N}^2$  such that  $w[i_1, i_2] = (a, a)$ . Then  $(i_1, i_2) \in Q$ , and since  $w_{i_1} = w_{i_2} = a$ , we also have that  $(i_1, i_1)$ ,  $(i_2, i_2)$  and  $(i_2, i_1)$  are in  $Q$ . We finally set  $S := \{i_1, i_2\}$ , which allows us to conclude.  $\square$

### 2.4.2 Thick closed subsets

In this subsection, we construct different examples of closed subsets of  $\beta(\mathbb{N}^2)$  in order to build an intuition on a condition that would characterize the points of  $X_2^{a,a}$ . Our study eventually leads us to formulate Conjecture 2.40. We start by observing that the two conditions that we introduced in Proposition 2.21 and Proposition 2.29 are insufficient to characterise  $X_2^{a,a}$ , as we show in the following example.

**Example 2.30.** Consider the family of subsets of  $\mathbb{N}^2$  obtained by removing finitely many points out of the diagonal,

$$\{\Delta \setminus F^2 : F \subseteq \mathbb{N} \text{ is finite}\}.$$

Note that this is a meet-semi lattice in  $\mathcal{P}(\mathbb{N}^2)$  that does not contain the empty set: the top element is  $\Delta = \Delta \setminus \emptyset$  and if  $F_1, F_2$  are two finite subsets of  $\mathbb{N}$ , then

$$(\Delta \setminus F_1^2) \cap (\Delta \setminus F_2^2) = \Delta \setminus (F_1 \cup F_2)^2,$$

where  $F_1 \cup F_2$  is still finite. Now, consider the closed subset of  $\beta(\mathbb{N}^2)$

$$C_{\Delta_{fin}} := \bigcap_{\substack{F \subseteq \mathbb{N} \\ F \text{ finite}}} \widehat{\Delta \setminus F^2}.$$

First, we prove that  $C_{\Delta_{fin}}$  satisfies the properties introduced in Proposition 2.21, and Proposition 2.29. For the first property, we start by observing that, for every finite subset  $F$  of  $\mathbb{N}$ , we have

$$R(\Delta \setminus F^2) = R\left(\bigcup_{i \notin F} \{(i, i)\}\right) = \Delta \setminus F^2.$$

Now, for all  $Q \subseteq \mathbb{N}^2$  such that  $C_{\Delta_{fin}}$  is contained in  $\widehat{Q}$ , there exists a finite subset  $F \subseteq \mathbb{N}$  such that

$$C_{\Delta_{fin}} \subseteq \widehat{\Delta \setminus F^2} \subseteq \widehat{Q}$$

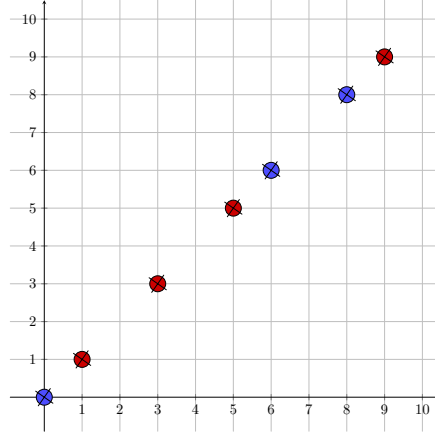
and since  $R$  is monotone,

$$\widehat{\Delta \setminus F^2} \subseteq \widehat{R(Q)}$$

which allows us to conclude. For the second property, let us pick an element

$$(n, m) \in \text{Cont}(C_{\Delta_{fin}}) = \bigcap_{\substack{F \subseteq \mathbb{N} \\ F \text{ finite}}} \Delta \setminus F^2.$$

Then,  $(n, m)$  belongs in particular to  $\Delta \setminus \{n, m\}^2$ , which is absurd. We conclude that  $\text{Cont}(C_{\Delta_{fin}})$  is the empty set, which is a set-theoretic square.

Figure 11:  $\Delta \setminus F$ ,  $Q_1$  and  $Q_2$ , with  $F = \{2, 4, 7, 10\}$ .

We now prove that  $C_{\Delta_{fin}}$  does not belong to  $X_2^{a,a}$  by using Proposition 2.19 (3). Denote by  $E$ , respectively  $O$ , the set of even, respectively odd, numbers in  $\mathbb{N}$ . Set  $Q_1 := \Delta \cap E^2$  and  $Q_2 := \Delta \cap O^2$ , which are disjoint. We know that  $C_{\Delta_{fin}} \subseteq \widehat{\Delta}$  and we notice that  $C_{\Delta_{fin}} \cap \widehat{Q_1}$  and  $C_{\Delta_{fin}} \cap \widehat{Q_2}$  are non-empty. A finite subset  $S \subseteq \mathbb{N}$  which has a non-empty intersection with both  $Q_1$  and  $Q_2$  has at least two elements, since  $Q_1$  and  $Q_2$  are disjoint. In particular, it cannot be contained in  $\Delta$ . By applying Proposition 2.19 (3), we conclude that  $C_{\Delta_{fin}}$  is not in  $X_2^{a,a}$ .

Building on this last example, we understand that, whenever  $C \in X_2^{a,a}$  is infinite, the clopen subsets  $\widehat{Q}$  which contain  $C$  all have to be "thick" around the diagonal, in a sense that we formalize here-below.

**Definition 2.31.** For any  $n \in \mathbb{N}$ , we set

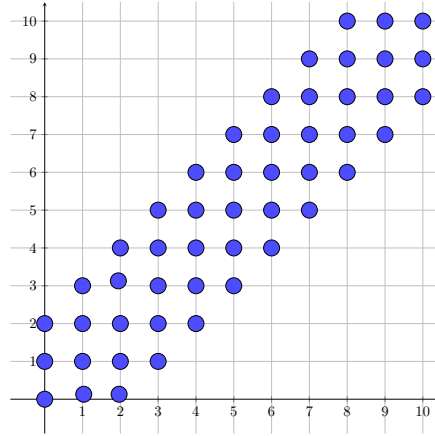
$$\Delta^n := \bigcup_{i \in \mathbb{N}} \{i + k : 0 \leq k \leq n\}^2.$$

In particular, for  $n = 0$ , we obtain the diagonal of  $\mathbb{N}^2$ ,  $\Delta^0 = \Delta$ .

We say that a clopen  $K$  of  $\beta(\mathbb{N}^2)$  is *thin* if, and only if, there exists  $n \in \mathbb{N}$  such that

$$K \cap \mathbb{N}^2 \subseteq \Delta^n.$$

We say that a clopen is *thick* if, and only if, it is not thin. More generally, we say that a closed subset  $C \in \mathcal{V}(\beta(\mathbb{N}^2))$  is thick if  $C$  is an intersection of thick clopen subsets.

Figure 12:  $\Delta^3$ .

**Example 2.32.** We review the closed subsets of  $\beta(\mathbb{N}^2)$  we previously mentioned under the light of Definition 2.31.

- A closed subset  $C$  contained in  $\mathbb{N}^2$  cannot be thick. Indeed,  $C = \widehat{Q}$  for some finite set  $Q$  of  $\mathbb{N}^2$ . Setting  $N := \max(\pi_1(Q), \pi_2(Q))$ , where  $\pi_1, \pi_2 : \mathbb{N}^2 \rightarrow \mathbb{N}$  are the canonical projections, we have that  $Q \subseteq \Delta^N$ .
- The remainder  $^*(\mathbb{N}^2)$  of  $\beta(\mathbb{N}^2)$  is thick. Indeed, as we proved in Lemma 1.15, a clopen subset of  $\beta(\mathbb{N}^2)$  contains  $^*(\mathbb{N}^2)$  if, and only if, its content is cofinite. Now, it is clear that a cofinite subset of  $\mathbb{N}^2$  is thick, which allows us to conclude.
- For any closed subset of  $\beta(\mathbb{N})$  that is not contained in  $\mathbb{N}$ , the closed subset  $C \otimes C$ , where  $\otimes$  has been introduced in Definition 1.27 is thick. Indeed, by definition of  $C \otimes C$ , a clopen subset of  $\beta(\mathbb{N})$  contains  $C \otimes C$  if, and only if, it is of the form  $\widehat{P^2}$ , where  $P \subseteq \mathbb{N}$  is such that  $C \subseteq \widehat{P}$ . In particular, since  $C$  is not contained in  $\mathbb{N}$ , such a subset  $P$  is necessarily infinite, which means that  $\widehat{P^2}$  is thick, and, a fortiori,  $C \otimes C$  too.
- The closed subset  $C_{\Delta_{fin}}$  introduced in Example  $C_{\Delta_{fin}}$  is not thick. Indeed, it is contained in  $\widehat{\Delta}$ , which is thin.

We shall prove that every infinite closed subset in  $X_2^{a,a}$  is thick.

**Proposition 2.33.** *A closed subset  $C \in X_2^{a,a}$  which is infinite is necessarily thick.*

*Proof.* Fix  $C \in X_2^{a,a}$  which is infinite. Let us assume, by absurd, that  $C$  is thin. Then, there exists  $n \in \mathbb{N}$  and  $Q \subseteq \mathbb{N}^2$  such that

$$C \subseteq \widehat{Q} \subseteq \widehat{\Delta^n}.$$

Since  $C$  is infinite, then by applying Proposition 1.19, we can construct a family of pairwise disjoint subsets  $Q_1, \dots, Q_{n+1}$  of  $\mathbb{N}^2$  such that, for every  $i \in \{1, \dots, n+1\}$ ,  $C \cap \widehat{Q_i}$

is non-empty. Since these subsets are pairwise disjoint, a finite subset  $S \subseteq \mathbb{N}$  such that  $S^2$  has a non-empty intersection with every  $Q_i$ , with  $i \in \{1, \dots, n+1\}$ , has at least  $n+1$  elements. In particular, it cannot be contained in  $Q$ , since  $Q \subseteq \Delta^n$ . Therefore, by applying Proposition 2.19 (3), we deduce that  $C$  is not in  $X_2^{a,a}$ , which is absurd: thus  $C$  is thick.  $\square$

The notion of thickness does not suffice in order to characterise  $X_2^{a,a}$ . We construct an example of a thick closed subset of  $\beta(\mathbb{N}^2)$ , which also satisfies the necessary conditions introduced in Proposition 2.21 and 2.29, and yet, does not belong to  $X_2^{a,a}$ .

**Example 2.34.** Consider the two closed subsets of  $\beta(\mathbb{N}^2)$

$$C_1 := \widehat{(2\mathbb{N})^2}$$

and

$$C_2 := \bigcap_{\substack{F \subseteq 2\mathbb{N}+1 \\ F \text{ finite}}} \widehat{\Delta_F^{2\mathbb{N}+1}}$$

where

$$\Delta_F^{2\mathbb{N}+1} = (\Delta \cap (2\mathbb{N}+1)^2) \setminus F^2,$$

and set

$$C := C_1 \cup C_2 \in \mathcal{V}(\beta(\mathbb{N}^2)).$$

By definition, for every  $Q \subseteq \mathbb{N}^2$ ,  $C \subseteq \widehat{Q}$  if, and only if,  $C_1 \subseteq \widehat{Q}$  and  $C_2 \subseteq \widehat{Q}$ . This is equivalent to saying that, for every  $Q \subseteq \mathbb{N}^2$ , there exists a finite subset  $F \subseteq 2\mathbb{N}+1$  such that

$$(2\mathbb{N})^2 \cup \Delta_F^{2\mathbb{N}+1} \subseteq Q.$$

We prove that  $C$  is thick, satisfies the conditions from Proposition 2.29 and Proposition 2.21, and yet, does not belong to  $X_2^{a,a}$ .

- We prove that  $C$  is thick. As we observed, every  $Q \subseteq \mathbb{N}^2$  such that  $\widehat{Q}$  contains  $C$  has to contain  $(2\mathbb{N})^2$ . In particular for every  $n \in \mathbb{N}$ ,  $\widehat{\Delta^n}$  cannot be contained in  $\widehat{Q}$ :  $C$  is thick.
- We prove that  $C$  satisfies the condition introduced in Proposition 2.29. Fix  $Q \subseteq \mathbb{N}^2$  such that  $\widehat{Q}$  contains  $C$ . As we mentioned, this is equivalent to saying that there exists a finite subset  $F \subseteq 2\mathbb{N}+1$  such that  $C \subseteq \widehat{(2\mathbb{N})^2 \cup \Delta_F^{2\mathbb{N}+1}} \subseteq \widehat{Q}$ . We want to prove that  $\widehat{R(Q)}$  contains  $C$ . First, observe that

$$R((2\mathbb{N})^2 \cup \Delta_F^{2\mathbb{N}+1}) = \bigcup_{S \subseteq 2\mathbb{N}} S^2 \cup \bigcup_{i \in 2\mathbb{N}+1 \setminus F} \{(i, i)\} = (2\mathbb{N})^2 \cup \Delta_F^{2\mathbb{N}+1}.$$

In particular, since  $R$  is monotone, we obtain that the finite subset  $F \subseteq 2\mathbb{N}+1$  is such that  $(2\mathbb{N})^2 \cup \Delta_F^{2\mathbb{N}+1} \subseteq R(Q)$ , which ends to prove that  $\widehat{R(Q)}$  contains  $C$ .

- We prove that  $C$  satisfies the condition introduced in Proposition 2.21. This is clear, since  $\text{Cont}(C) = (C_1 \cap \mathbb{N}^2) \cup (C_2 \cap \mathbb{N}^2) = (2\mathbb{N})^2$ .

We now prove that  $C \notin X_2^{a,a}$ . Fix a two colours colouring  $(P_1, P_2)$  of  $2\mathbb{N} + 1$  and set

$$\begin{cases} Q := (2\mathbb{N})^2 \cup ((2\mathbb{N} + 1)^2 \cap \Delta) \\ Q_1 := P_1^2 \cap \Delta \\ Q_2 := P_2^2 \cap \Delta \end{cases}$$

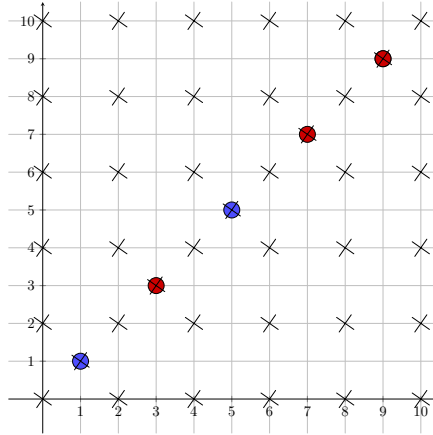


Figure 13:  $Q$ ,  $Q_1$  and  $Q_2$ .

We have that  $C \subseteq \widehat{Q}$ , and  $C$  has a non-empty intersection with both  $\widehat{Q}_1$  and  $\widehat{Q}_2$ . A subset of  $\mathbb{N}^2$  intersecting both  $Q_1$  and  $Q_2$  needs to contain an element  $(n, m)$ , with  $n \neq m$  and  $n, m \in 2\mathbb{N} + 1$ . In particular, it cannot be contained in  $Q$ . We deduce that there exists no finite subset  $S \subseteq \mathbb{N}$  such that  $S^2$  is contained in  $Q$  and has a non-empty intersection with both  $Q_1$  and  $Q_2$ . We conclude, by Proposition 2.19 (3) that  $C$  does not belong to  $X_2^{a,a}$ .

From Example 2.34, we derive the following insight. For any closed subset  $C \in X_2^{a,a}$  that is not included in  $\mathbb{N}^2$ , if we fix a subset  $Q$  of  $\mathbb{N}^2$  such that  $\widehat{Q}$  contains  $C$ , and an integer  $n \in \mathbb{N}$ , we need to pay a particular attention to the finite set-theoretic squares  $S^2$  contained in  $Q \cap \Delta^n$ .

**Definition 2.35.** For any  $Q \subseteq \mathbb{N}^2$  and for any  $n \in \mathbb{N}$ , we set

$$\mathcal{S}_{Q, \Delta^n} := \{S \in \mathcal{S}_Q : S^2 \subseteq \Delta^n\},$$

where  $\mathcal{S}_Q$  has been introduced in Definition 2.27. In particular,

$$R(Q) = \bigcup_{n \in \mathbb{N}} \bigcup_{\mathcal{S}_{Q, \Delta^n}} S^2.$$

For a fixed  $Q \subseteq \mathbb{N}^2$ , and a fixed  $n \in \mathbb{N}$ , we now introduce the two following subsets of  $\mathcal{S}_{Q, \Delta^n}$ . First, we define  $\mathcal{T}_{Q, n}$ , consisting of the finite subsets  $S \subseteq \mathbb{N}$  such that  $S^2$  is contained in  $Q$ , and there exists a finite square  $T_S^2$  contained in  $Q$  which contains  $S^2$  and is not contained in  $\Delta^n$

$$\mathcal{T}_{Q, n} := \{S \in \mathcal{S}_{Q, \Delta^n} : \exists T_S^2 \subseteq Q \text{ such that } S \subseteq T_S \text{ and } T_S^2 \not\subseteq \Delta^n\}.$$

Then, we define  $\mathcal{U}_{Q, n}$  consisting of the finite subsets  $S \subseteq \mathbb{N}$  such that  $S^2$  is contained in  $Q$ , and every finite square  $T_S^2$  contained in  $Q$  which contains  $S^2$  has to be contained in  $\Delta^n$ ,

$$\mathcal{U}_{Q, n} := \mathcal{S}_{Q, \Delta^n} \setminus \mathcal{T}_{Q, n} = \{S \in \mathcal{S}_{Q, \Delta^n} : \forall T_S^2 \subseteq Q : (S \subseteq T_S \Rightarrow T_S^2 \subseteq \Delta^n)\}.$$

**Example 2.36.** For more clarity, we illustrate Definition 2.35 on a simple example. Consider the subset of  $\mathbb{N}^2$

$$Q = \{(1, 1)\} \cup \{(2, 2)\} \cup \{(1, 4)\} \cup \{(4, 1)\} \cup (\mathbb{N} \setminus \{1, 2\})^2.$$

Notice that this is an union of finite set-theoretic squares, and thus  $R(Q) = Q$ . By applying Definition 2.35, we obtain that

- $\mathcal{S}_{Q, \Delta^0} = \{\{i\} : i \geq 1\}$ ;
- $\mathcal{S}_{Q, \Delta^1} = \{\{1\}, \{2\}\} \cup \{\{i, i+1\} : i \geq 3\}$ ;
- $\mathcal{S}_{Q, \Delta^3} = \{\{1\}, \{2\}, \{1, 4\}\} \cup \{\{i, i+1, i+2, i+3\} : i \geq 3\}$ .

Also,

- $\mathcal{T}_{Q, \Delta^0} = \{\{i\} : i \neq 2\}$  and  $\mathcal{U}_{Q, \Delta^0} = \{\{2\}\}$ ;
- $\mathcal{T}_{Q, \Delta^1} = \{\{1\}\} \cup \{\{i, i+1\} : i \geq 3\}$  and  $\mathcal{U}_{Q, \Delta^1} = \{\{2\}\}$ ;
- $\mathcal{T}_{Q, \Delta^3} = \{\{i, i+1, i+2, i+3\} : i \geq 3\}$  and  $\mathcal{U}_{Q, \Delta^3} = \{\{1\}, \{2\}, \{1, 4\}\}$ .

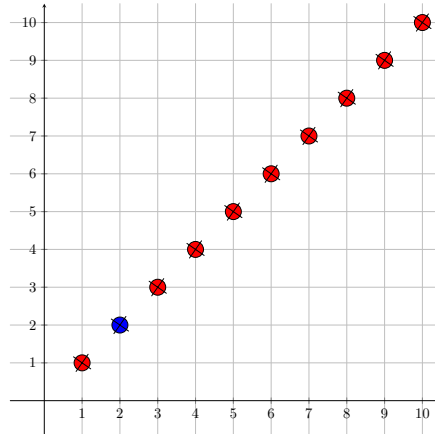


Figure 14:  $\bigcup_{S \in \mathcal{S}_{Q, \Delta^0}} S^2$ ,  $\bigcup_{S \in \mathcal{T}_{Q, 0}} S^2$  and  $\bigcup_{S \in \mathcal{U}_{Q, 0}} S^2$ .

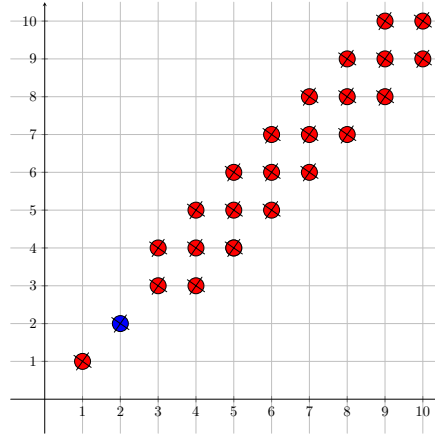


Figure 15:  $\bigcup_{S \in S_{Q, \Delta^1}} S^2$ ,  $\bigcup_{S \in T_{Q, 1}} S^2$  and  $\bigcup_{S \in U_{Q, 1}} S^2$ .

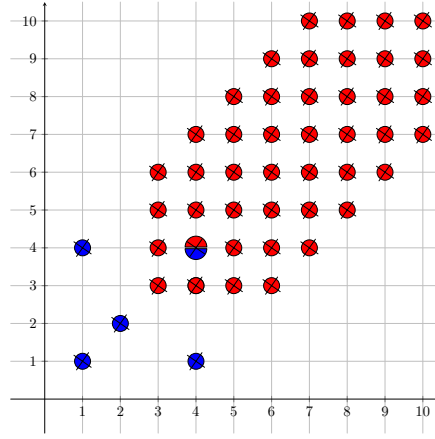


Figure 16:  $\bigcup_{S \in S_{Q, \Delta^3}} S^2$ ,  $\bigcup_{S \in T_{Q, 3}} S^2$  and  $\bigcup_{S \in U_{Q, 3}} S^2$ .



With these notations in mind, we prove the following statement, which can be understood as a refinement of Proposition 2.29. Considering a clopen  $\widehat{Q}$  of  $\beta(\mathbb{N}^2)$  which contains  $C$ , it allows us to obtain a clopen contained in  $\widehat{R(Q)}$  which still contains  $C$  by removing certain infinite families of finite squares out of  $R(Q)$ .

**Definition 2.37.** For any  $Q \subseteq \mathbb{N}^2$  and any  $n \in \mathbb{N}$ , we consider the following subsets of  $R(Q)$ :

$$R_{\text{bound}}^n(Q) := \{\bar{p} \in \mathbb{N}^2 : \exists S \in \mathcal{U}_{Q,n}, \bar{p} \in S^2\}$$

and

$$R_{\text{unb}}^n(Q) = R(Q) \setminus R_{\text{bound}}^n(Q).$$

**Proposition 2.38.** Fix an infinite  $C \in X_2^{a,a}$  and fix  $Q \subseteq \mathbb{N}^2$  such that  $\widehat{Q}$  contains  $C$ . Then, for every  $n \in \mathbb{N}$ ,

$$C \cap \widehat{R_{\text{bound}}^n(Q)} = \emptyset,$$

and, in particular, we have the series of inclusions

$$C \subseteq \widehat{R_{\text{unb}}^n(Q)} \subseteq \widehat{R(Q)} \subseteq \widehat{Q}.$$

*Proof.* Consider an infinite closed subset  $C \in X_2^{a,a}$ . Fix  $Q \subseteq \mathbb{N}^2$  such that  $\widehat{Q}$  contains  $C$ , and assume, by absurd, that there exists  $n \in \mathbb{N}$  such that

$$C \cap \widehat{R_{\text{bound}}^n(Q)} \neq \emptyset,$$

We prove that  $C$  cannot satisfy Proposition 2.19 (3). First, recall that, by proposition 2.29,  $C$  is contained in  $\widehat{R(Q)}$ . Second, recall that, we proved in Proposition 2.33 that  $C$  is thick, and thus  $C$  has a non-empty intersection with  $\mathbb{N}^2 \setminus \Delta^n$ . Note that  $\mathbb{N}^2 \setminus \Delta^n$  and  $R_{\text{bound}}^n(Q)$  are disjoint.

Now, applying Proposition 2.19 (3), there should exist a finite subset  $S \subseteq \mathbb{N}$  such that  $S^2 \subseteq R(Q)$ ,  $S^2 \not\subseteq \Delta^n$  and  $S^2 \cap R_{\text{bound}}^n(Q)$  is non-empty. Pick  $(i, j)$  in this intersection. By definition of  $R_{\text{bound}}^n(Q)$ ,  $(i, j)$  cannot be contained in a finite set-theoretic square which is contained in  $Q$  but not contained in  $\Delta^n$ . Since  $S^2$  is exactly a subset, of this form, and that  $(i, j) \in S^2$ , we obtain the desired contradiction. We conclude that, for every  $n \in \mathbb{N}$ ,

$$C \cap \widehat{R_{\text{bound}}^n(Q)} = \emptyset,$$

and since  $C \subseteq \widehat{R(Q)} \subseteq \widehat{Q}$ , we conclude that

$$C \subseteq \widehat{R_{\text{unb}}^n(Q)} \subseteq \widehat{R(Q)} \subseteq \widehat{Q}.$$

□

**Example 2.39.** Fix a closed subset  $C_a \in \mathcal{V}(\beta(\mathbb{N}))$ , and set  $C_{a,a} := C_a \otimes C_a$  as introduced in Definition 1.27. We prove that  $C_{a,a}$  is in  $X_2^{a,a}$  by applying Proposition 2.19 (3). Fix a

finite family of subsets  $Q, Q_1, \dots, Q_\ell$  of  $\mathbb{N}^2$ , such that  $C_{a,a} \subseteq \widehat{Q}$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $C_{a,a} \cap Q_i$  is non-empty. By definition of  $C_{a,a}$ , there exists a subset  $P \subseteq \mathbb{N}$ , with  $C_a \subseteq \widehat{P}$ , such that

$$C_{a,a} \subseteq \widehat{P^2} \subseteq \widehat{Q}.$$

In particular, for every  $i \in \{1, \dots, \ell\}$ ,  $P^2 \cap Q_i$  is non-empty: pick  $(n_i, m_i)$  in this intersection. The finite subset  $S := \{n_i, m_i : i \in \{1, \dots, \ell\}\}$  is, by construction, such that  $S^2 \subseteq Q$  and, for every  $i \in \{1, \dots, \ell\}$ ,  $S^2 \cap Q_i$  is non-empty. By Proposition 2.19 (3), we conclude that  $C_{a,a}$  is in  $X_2^{a,a}$ .

This last example allows us to formulate the following conjecture: the infinite closed subsets in  $X_2^{a,a}$  consists exactly in the closed subsets of  $\beta(\mathbb{N}^2)$  which can be written as a product of an infinite closed subset of  $\beta(\mathbb{N})$  with itself.

**Conjecture 2.40.** *An infinite closed subset  $C_{a,a} \in \mathcal{V}(\beta(\mathbb{N}^2))$  is in  $X_2^{a,a}$  if, and only if, there exists an infinite closed subset  $C_a \in \mathcal{V}(\beta(\mathbb{N}))$  such that  $C_{a,a} = C_a \otimes C_a$ .*

## Conclusion of the chapter

In this chapter, we provided a detailed duality theoretic treatment of the Boolean algebra of languages corresponding to Boolean combinations of sentences written by using a block of  $k$  existential quantifiers, letter predicates, and uniform numerical predicates. Several directions can be taken in order to continue this work.

First, pursuing the goal of Section 2.4, we still lack a necessary and sufficient condition which would allow for an explicit description of  $X_2^{a,a}$ . This could be use to obtain a different characterisation of  $X_2$ , and, likely,  $X_k$ , which could deepen the understanding we obtained on it via Proposition 2.14. A natural thing to start with would be checking whether Conjecture 2.40 is true.

Second, note that, in this thesis, we restricted our attention to subfragments of the fragment  $\mathcal{BS}[\mathcal{N}]$ . This last fragment has been well studied in the case of regular languages, but still admits no satisfying duality-theoretic treatment. Also, it would be interesting to understand the duality theory for Boolean algebras obtained by applying a layer of existential quantifier to an arbitrary subalgebra of the Boolean algebra corresponding to quantifier-free formulas with  $k$ -ary uniform numerical predicates, and letter predicates.

Another question that could be asked, and which it is necessary to solve in order to understand  $FO[\mathcal{N}]$ , is how is it possible to take into account alternation of quantifiers. The main ingredient that allows us to conduct our study in this thesis is the fact that a block of existential quantifier commutes with finite disjunctions. However, it is not clear how one can rely on our work to think about alternation of universal and existential quantifiers. Understanding mixed steps of universal and existential quantification is

a complex problem. We refer to [11] for an approach that tries to aim toward that direction, and is based on substitution of formulas and transductions.

Finally, we made the assumption at the beginning that all of the numerical predicates we consider are uniform. One could try to understand whether it would be possible to build on our framework in order to think about arbitrary numerical predicates.

---

## Ultrafilter equations for the existential fragment of logic on words

---

In this chapter, we treat the topic of ultrafilter equations for the existential fragment of first-order logic on words. Fix an integer  $k \geq 1$ . In the previous chapter, we introduced  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ , a Boolean algebra corresponding to a certain fragment of logic on words. This fragment corresponds to Boolean combinations of sentences defined by using a block of at most  $k$  existential quantifiers, letter predicates and uniform numerical predicates of arity  $l \in \{1, \dots, k\}$ . We gave different characterisations of the dual space associated to this Boolean algebra, see Proposition 2.7 and Proposition 2.14. As we mentioned previously, there are few instances of computations of dual spaces corresponding to fragments of logic defined by languages that lie outside of the scope of regular languages. There are even less instances of computations of a family of ultrafilter equations which characterise a Boolean algebra of this kind. A question we want to investigate the answer for in this thesis is whether one could obtain a good description in terms of ultrafilter equations of the existential fragment of first-order logic on words

$$\mathcal{B}\Sigma_1[\mathcal{N}] := \bigcup_{k \in \mathbb{N}} \mathcal{B}\Sigma_1[\mathcal{N}_k],$$

defined by considering numerical predicates of arbitrary arity. Recall that nullary numerical predicates are all uniform. An important first step that should be made is to obtain and understand ultrafilter equations for  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_k^u]$  and  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ , for any  $k \geq 1$ . An explicit basis of ultrafilter equations for  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_1^u]$  is already available in [29]. However, the reasoning conducted there does not directly allow for a generalization for the case  $k \geq 2$ , and therefore an extra step has to be performed in order to understand the general case. Our contribution is a more topological presentation of the ultrafilter equations for  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_1^u]$ . We reduce the combinatorics that was involved in the reasonings used to prove soundness and completeness in [29] to a bare minimum. In

particular, our approach allows us to exhibit a basis of ultrafilter equations for  $\mathcal{B}\Sigma_1[\mathcal{N}_1^u]$ . The main ingredient we use from the previous chapter is the approach introduced in Section 2.3, which involves finite colourings of  $\mathbb{N}$ . This allows for a reformulation of the ultrafilter equations from [29] in terms of the existence of a finite colouring of  $\mathbb{N}$  which satisfies certain properties.

**Outline of the chapter:** In Section 3.1, we give a different presentation of the Boolean algebras  $\mathcal{B}_k$  and  $\mathcal{B}_{0,k}$ , for any  $k \geq 1$ , taking inspiration from the characterisation of the points of the dual space  $X_k$  we provided in Proposition 2.14. In Section 3.2, we describe a general family of ultrafilter equations, which will encompass every equation needed in order to describe  $\mathcal{B}_1$  and  $\mathcal{B}_{0,1}$ . We then give a reformulation of this general family of equations in terms of a condition relative to a finite colourings of  $\mathbb{N}$ . This will simplify the reasoning which will follow, and in Section 3.3 we use this reformulation in order to show soundness and completeness for the equations we introduced to describe  $\mathcal{B}_1$  and  $\mathcal{B}_{0,1}$ .

### 3.1 An alternative presentation for $\mathcal{B}_k$ and $\mathcal{B}_{0,k}$

In this section, we give a different presentation of the Boolean algebras  $\mathcal{B}_k$  and  $\mathcal{B}_{0,k}$ , for any  $k \geq 1$ , taking inspiration from the characterisation of the points of the dual space  $X_k$  we provided in Proposition 2.14. This presentation, in terms of languages associated to finite colourings of  $\mathbb{N}^k$ , will greatly simplify the proofs of soundness and completeness which will follow, and allow for an enlightening reformulation of the ultrafilter equations we will consider.

Let us explain this setting in the case where  $k = 1$ . Let us look again at the languages  $L_{\diamond_Q^a}$ , where  $a \in A$  and  $Q \subseteq \mathbb{N}$ , introduced at the very beginning of Section 2.1. Instead of fixing a subset  $Q$  of  $\mathbb{N}$ , a letter  $a \in A$ , and considering the set of all words such that there exists a position  $i < |w|$  such that  $w_i = a$ , we could rather fix a finite colouring of  $\mathbb{N}$ , and consider the words such that the content of  $w$  on each colour is exactly a given subset of  $A$ . The languages that we obtain this way are related to the notion of profile that we introduced in Definition 2.12, and we prove that they allow for an alternative description of the Boolean algebra  $\mathcal{B}_1$  and  $\mathcal{B}_{0,1}$ . This idea can be generalized for any  $k \geq 1$ , and this motivates the introduction of the languages that we define here-below.

In the rest of the chapter, for any  $k \geq 1$ , to mention a finite colouring of  $\mathbb{N}^k$  with  $\ell \geq 1$  colours, we use the notation  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$ .

**Definition 3.1.** For any  $Q \subseteq \mathbb{N}^k$  and any  $B \subseteq A^k$ , we consider the language  $K_{Q,B}$  of all words having content  $B$  on  $Q$ ,

$$K_{Q,B} := \{w \in A^* : \langle w, Q \rangle = B\}.$$

More generally, for any  $k, \ell \geq 1$ , any finite colouring  $\mathcal{Q}$  of  $\mathbb{N}^k$  with  $\ell$  colours, and for

any family  $\bar{B} = (B_1, \dots, B_\ell)$  of  $\ell$  subsets of  $A^k$ , we consider the language of the words having content on  $Q_j$  equal to  $B_j$ , for every  $j \in \{1, \dots, \ell\}$ ,

$$K_{Q, \bar{B}} := \bigcap_{j=1}^{\ell} K_{Q_j, B_j}.$$

Observe that these languages can be seen as the equivalence classes for a certain equivalence relation on  $A^*$ . Indeed, for any  $Q \subseteq \mathbb{N}^k$ , we set

$$\sim_Q := \{(w_1, w_2) \in (A^*)^2 : \langle w_1, Q \rangle = \langle w_2, Q \rangle\}.$$

Now, for any finite word  $w_1 \in A^*$ , setting  $B := \langle w_1, Q \rangle$ , the equivalence class which contains  $w_1$  is

$$[w_1]_{\sim_Q} = \{w_2 \in A^* : \langle w_1, Q \rangle = \langle w_2, Q \rangle\} = \{w_2 \in A^* : \langle w_1, Q \rangle = B\} = K_{Q, B}.$$

Note that this equivalence relation is finitely indexed, since  $\mathcal{P}(A^k)$  is finite.

More generally, for any  $k, \ell \geq 1$ , and any finite colouring of  $\mathbb{N}^k$  with  $\ell$  colours  $Q$ , we set

$$\sim_Q := \bigcap_{j=1}^{\ell} \sim_{Q_j},$$

and an equivalence class for this relation corresponds to a language of the form  $K_{Q, \bar{B}}$ , for some  $\bar{B} \in \mathcal{P}(A^k)^\ell$ . This equivalence relation is also finitely indexed, since  $\mathcal{P}(A^k)^\ell$  is finite. These languages allow for a reformulation of the generators of the Boolean algebra  $\mathcal{B}_k$ , for any  $k \geq 1$ , which will greatly simplify our considerations in the upcoming sections.

**Proposition 3.2.** *For any  $k \geq 1$ , the Boolean algebra  $\mathcal{B}_k$  is generated by the languages  $K_{Q, \bar{B}}$ , where  $Q$  ranges over finite colourings of  $\mathbb{N}^k$  with  $\ell \geq 1$  colours and  $\bar{B}$  ranges over  $\mathcal{P}(A^k)^\ell$ .*

*Proof.* We know, by Proposition 2.5 that  $\mathcal{B}_k$  is generated by the languages  $L_{\diamond \bar{a}_Q}$  where  $Q$  ranges over subsets of  $\mathbb{N}^k$  and  $\bar{a}$  ranges over  $A^k$ . First, we prove that these languages can be expressed as a Boolean combination of languages of the form  $K_{Q, \bar{B}}$ . Let us consider the two colours colouring  $(Q, Q^c)$ . A finite word  $w \in A^*$  is in  $L_{\diamond \bar{a}_Q}$  if, and only if, the content of  $w$  on  $Q$  does contain the  $k$ -tuple  $\bar{a}$ , which allows us to write

$$L_{\diamond \bar{a}_Q} = \bigcup_{\substack{\bar{a} \in B \subseteq A^k \\ B' \subseteq A^k}} (K_{Q, B} \cap K_{Q^c, B'}) = \bigcup_{\substack{\bar{a} \in B \subseteq A^k \\ B' \subseteq A^k}} K_{(Q, Q^c), (B, B')}$$

and allows us to conclude.

Now, we fix a finite colouring  $Q$  of  $\mathbb{N}^k$  with  $\ell \geq 1$  colours and  $\bar{B} \in \mathcal{P}(A^k)^\ell$ , and we prove that  $K_{Q, \bar{B}}$  can be written as a Boolean combination of languages of the form  $L_{\diamond \bar{a}_Q}$ .

First, note that, since

$$K_{Q, \bar{B}} = \bigcap_{j=1}^{\ell} K_{Q_j, B_j},$$

we only need to prove the result for any  $K_{Q, B}$ , with  $Q \subseteq \mathbb{N}^k$  and  $B \subseteq A^k$ . Now, we have

$$\begin{aligned} K_{Q, B} &= \{w \in A^* : \langle w, Q \rangle = B\} \\ &= \{w \in A^* : \{\bar{a} \in A^k : c_{\bar{a}}(w) \cap Q \neq \emptyset\} = B\} \\ &= \left( \bigcap_{\bar{a} \in B} \{w \in A^* : c_{\bar{a}}(w) \cap Q \neq \emptyset\} \right) \cap \left( \bigcap_{\bar{a} \notin B} \{w \in A^* : c_{\bar{a}}(w) \cap Q = \emptyset\} \right) \\ &= \bigcap_{\bar{a} \in B} L_{\diamond_{\bar{a}} Q} \cap \bigcap_{\bar{a} \notin B} (L_{\diamond_{\bar{a}} Q})^c. \end{aligned}$$

which allows us to conclude.  $\square$

We give the analogous result for  $B_{0, k}$ , for any  $k \geq 1$ .

**Corollary 3.3.** *For any  $k \geq 1$ , the Boolean algebra  $B_{0, k}$  is generated by the languages  $L_P$  and  $K_{Q, \bar{B}}$  where  $P$  ranges over subsets of  $\mathbb{N}$ ,  $Q$  ranges over finite colourings of  $\mathbb{N}^k$  with  $\ell \geq 1$  colours and  $\bar{B}$  ranges over  $\mathcal{P}(A^k)^\ell$ .*

*Proof.* Direct consequence of Proposition 3.2 and Corollary 2.6.  $\square$

## 3.2 A certain family of ultrafilter equations

In this section, we introduce a general family of ultrafilter equations on  $\beta(A^*)$  which will encompass every ultrafilter equation we will require in order to describe  $\mathcal{B}_1$  and  $\mathcal{B}_{0, 1}$ . We then explain how that it is possible to reformulate these equations into a condition that requires the existence of a certain finite colourings of  $\mathbb{N}$ . This property will be our main tool in order to check soundness and completeness in Section 3.3.

### 3.2.1 General setting

Let us start by defining the family of ultrafilter equations on  $\beta(A^*)$  which will be at the center of our study. For any  $k, n \geq 1$  we use the notation  $A^* \otimes (\mathbb{N}^k)^n$  in order to refer to  $A^* \otimes \underbrace{(\mathbb{N}^k \times \dots \times \mathbb{N}^k)}_{n \text{ times}}$  introduced in Section 1.4.2.

**Definition 3.4.** For any  $k, n \geq 1$ , any finite family of maps  $p_1, \dots, p_n : A^* \otimes (\mathbb{N}^k)^n \rightarrow \mathbb{N}^k$  and  $u, v : A^* \otimes (\mathbb{N}^k)^n \rightarrow A^*$ , we denote by  $\mathcal{E}_{u=v}^{p_1, \dots, p_n}$  the family of ultrafilter equations

$$\beta u(\nu) \leftrightarrow \beta v(\nu),$$

where  $\nu$  ranges over all elements of  $\beta(A^* \otimes (\mathbb{N}^k)^n)$  such that

$$\beta p_1(\nu) = \dots = \beta p_n(\nu).$$

The ultrafilter equations we use in order to describe  $\mathcal{B}_1$  and  $\mathcal{B}_{0,1}$  are all particular instances of the ones introduced in Definition 3.4. Let us provide some intuition behind these equations. The Boolean algebra  $\mathcal{B}_{0,1}$  corresponds to  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_1^u]$ , the Boolean algebra of languages corresponding to Boolean combinations of sentences written by using unary uniform numerical predicates, and letter predicates. It has been proven in [29], Theorem 5.16, that  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_1^u] \cap \text{Reg}$ , the Boolean algebra of regular languages in  $\mathcal{B}\Sigma_1[\mathcal{N}_0, \mathcal{N}_1^u]$  is described by the profinite equations

$$(x^{\omega-1}s)(x^{\omega-1}t) = (x^{\omega-1}t)(x^{\omega-1}s) \text{ and } (x^{\omega-1}s)^2 = x^{\omega-1}s,$$

for  $x, s, t$  words of the same length. The profinite monoid on  $A^*$  is a compactification of  $A^*$  which embeds in  $\beta(A^*)$ , and thus the ultrafilter equations we want to obtain are, in a sense, a generalization of these profinite equations to the setting of ultrafilters.

In general, for any function  $f : X \rightarrow K$  from a set  $X$  to a compact Hausdorff space  $K$ , there exists a unique  $\beta f : \beta X \rightarrow K$  that is continuous and extends  $f$ . Therefore, in order to define any such continuous function on  $\beta X$ , it suffices to just define it on  $X$ .

For any  $k \geq 1$ , any finite word  $w \in A^*$ , any  $k$ -tuple of letters  $\bar{a} \in A^k$  and any family of distinct integers  $\bar{j} \in |w|^k$ , we define

$$w(\bar{j} \rightarrow \bar{a})$$

to be the word obtained by replacing, for any  $m \in \{1, \dots, k\}$ ,  $w_{j_m}$  by  $a_m$  in  $w$ . This allows us to define the map  $f_{\bar{a}} : A^* \otimes \mathbb{N}^k \rightarrow A^*$  as follows: for any  $(w, \bar{j})$  in  $A^* \otimes \mathbb{N}^k$ ,

$$f_{\bar{a}}(w, \bar{j}) := \begin{cases} w(\bar{j} \rightarrow \bar{a}) & \text{if all of the } j_m, \text{ for } m \in \{1, \dots, k\}, \text{ are distinct} \\ w & \text{otherwise} \end{cases},$$

and its continuous extension  $\beta f_{\bar{a}} : \beta(A^* \otimes \mathbb{N}^k) \rightarrow \beta(A^*)$ . Following the ideas introduced in Section 2.3, we see elements of  $\beta(\mathbb{N}^k)$  as generalized  $k$ -tuples of position. In particular, it is not equivalent to consider a generalized word with a  $k$ -tuple of positions, that is, an element of  $\beta(A^* \otimes \mathbb{N}^k)$ , and a generalized word with a generalized  $k$ -tuple of positions, that is, an element of  $\beta(A^*) \times \beta(\mathbb{N}^k)$ . If we want to consider the generalized  $k$ -tuple of positions associated to  $\nu \in \beta(A^* \otimes \mathbb{N}^k)$ , we look at the ultrafilter  $\beta\pi(\nu) \in \beta(\mathbb{N}^k)$ , where  $\pi : A^* \otimes \mathbb{N}^k \rightarrow \mathbb{N}^k$  is the canonical projection.

We now introduce the generalization of the profinite equation  $(x^{\omega-1}s)(x^{\omega-1}t) = (x^{\omega-1}t)(x^{\omega-1}s)$ , for  $x, s, t$  words of the same length, to the ultrafilter setting.

**Definition 3.5.** For any  $a, b \in A$ , we consider the map  $f_{a,b} : A^* \otimes \mathbb{N}^2 \rightarrow A^*$ , which sends any  $(w, j_1, j_2) \in A^* \otimes \mathbb{N}^2$  to

$$f_{a,b}(w, j_1, j_2) := \begin{cases} w((j_1, j_2) \rightarrow (a, b)) & \text{if } j_1 \neq j_2 \\ w & \text{otherwise} \end{cases}.$$



We denote by  $\pi_1, \pi_2 : A^* \otimes \mathbb{N}^2 \rightarrow \mathbb{N}$  the canonical projections on the first and second coordinate. Finally, for any  $a, b \in A$ , we denote by  $\mathcal{E}_{ab=ba}$  the family of ultrafilter equations  $\mathcal{E}_{f_{a,b}, f_{b,a}}^{\pi_1, \pi_2}$  as in Definition 3.4.

The following equations will be required in order to describe  $\mathcal{B}_{0,1}$ .

**Definition 3.6.** For any  $a, b \in A$ , we consider the map  $f_{a,a,b} : A^* \otimes \mathbb{N}^3 \rightarrow A^*$ , which sends any  $(w, j_1, j_2, j_3) \in A^* \otimes \mathbb{N}^3$  to

$$f_{a,a,b}(w, j_1, j_2, j_3) := \begin{cases} w((j_1, j_2, j_3) \rightarrow (a, a, b)) & \text{if } j_1 \neq j_2 \neq j_3 \\ w & \text{otherwise} \end{cases}.$$

We denote by  $\pi_1, \pi_2, \pi_3 : A^* \otimes \mathbb{N}^3 \rightarrow \mathbb{N}$  the canonical projections on the first, second and third coordinate. Finally, for any  $a, b \in A$ , we denote by  $\mathcal{E}_{aab=abb}$  the family of ultrafilter equations  $\mathcal{E}_{f_{a,a,b}, f_{a,b,b}}^{\pi_1, \pi_2, \pi_3}$  as in Definition 3.4.

Finally, we will need this last set of equations in order to describe  $\mathcal{B}_1$ .

**Definition 3.7.** For any  $a \in A$ , we consider the map  $f_a : A^* \otimes \mathbb{N} \rightarrow A^*$  which sends any finite word with a marked position  $(w, i)$  to  $w(i \rightarrow a)$ , and the map  $f_{a,a} : A^* \otimes \mathbb{N} \rightarrow A^*$  which sends any finite word with a marked position  $(w, i)$  to  $w(i \rightarrow a).a$ .

We denote by  $\pi : A^* \otimes \mathbb{N} \rightarrow \mathbb{N}$  the canonical projection and  $|\cdot| : A^* \otimes \mathbb{N} \rightarrow \mathbb{N}$  the map that sends  $(w, i) \in A^* \otimes \mathbb{N}$  to  $|w|$ . Finally, for any  $a \in A$ , we denote by  $\mathcal{E}_{a=a.a}$  the family of ultrafilter equations  $\mathcal{E}_{f_a, f_{a,a}}^{\pi, |\cdot|}$  as in Definition 3.4.

Our goal in the upcoming sections will be to prove the following results.

**Theorem 3.8** ([29], Theorem 4.7). *A language  $L \subseteq A^*$  is in  $\mathcal{B}_{0,1}$  if, and only if,  $L$  satisfies the families of ultrafilter equations  $\mathcal{E}_{ab=ba}$ , and  $\mathcal{E}_{aab=abb}$ , for every  $a, b \in A$ .*

**Theorem 3.9.** *A language  $L \subseteq A^*$  is in  $\mathcal{B}_1$  if, and only if,  $L$  satisfies the families of ultrafilter equations  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  and  $\mathcal{E}_{a=a.a}$ , for every  $a, b \in A$ .*

The first of these two results has been proven in [29]: we briefly describe the approach that was taken. Let us focus on the proof of completeness, as it is the most difficult part of the problem. Fix a language  $L \subseteq A^*$ . A permutation  $\sigma$  with finite support is said to be *compatible* with  $L$  provided that, for all  $w \in A^*$ , if the support of  $\sigma$  is contained in  $\{0, \dots, |w| - 1\}$ , then  $w$  is in  $L$  if, and only if, the word obtained by applying  $\sigma$  to the set of positions of  $w$ , and permutting the letters of  $w$  accordingly, is in  $L$ . We now define a binary relation  $R_L$  on  $\mathbb{N}$  as follows:  $i R_L j$  if, and only if, the transposition  $(ij)$  is compatible with  $L$ . They observe that, if  $L$  satisfies the family of ultrafilter equations  $(\mathcal{E}_{ab=ba})_{a,b \in A}$ , then  $R_L$  contains an equivalence relation of finite index. Let  $\theta$  be an equivalence relation of finite index contained in  $R_L$ . They prove that, if  $L$  also satisfies the family of ultrafilter equations  $(\mathcal{E}_{aab=abb})_{a,b \in A}$ , then there exists  $n \in \mathbb{N}$  such that for all  $u, v \in A^*$ , if  $|u| = |v|$ , and for each  $\theta$ -equivalence class  $P$ ,  $u \sim_P v$  (see Definition

3.1 ) then  $u$  is in  $L$  if, and only if,  $v$  is in  $L$ . The proof of this statement involves thorough reasonings about finite structures. This defines an equivalence relation on  $A^*$ , and they conclude the proof by showing that  $L$  can be written as a Boolean combination of languages defined by using some of the equivalence classes for this relation.

Our approach for the proof is different, in the sense that it keeps the combinatoric reasonings to a minimum. We replace the intermediate steps from [29] that we just described by a complete reformulation of the ultrafilter equation. By using finite colourings, we are able to obtain a characterisation where ultrafilters do not intervene anymore: this allows to reason, most of the time, with finite words instead of ultrafilters on words. The advantage is two-fold: first, we can more easily apply the intuition we have on the problem from finite words, and second, we avoid the difficulties that can happen whenever conducting proofs relative to ultrafilters. This reformulation was sufficient for us to build enough intuition to come up with Theorem 3.9, and giving a straight-forward proof of both completeness and soundness.

### 3.2.2 Ultrafilter equations in terms of finite colourings

In this subsection, we present a reformulation of the equations introduced in Definition 3.4 in terms of finite colourings. First, we prove a technical lemma that allows for a rephrasing of one of the conditions involved in Definition 3.4.

**Lemma 3.10.** *For any  $k, n \geq 1$ , consider an ultrafilter  $\nu \in \beta(A^* \otimes (\mathbb{N}^k)^n)$  and a family of  $n$  maps  $p_1, \dots, p_n : A^* \otimes (\mathbb{N}^k)^n \rightarrow \mathbb{N}^k$ . For any  $\alpha \in \beta(\mathbb{N}^k)$ , the following statements are equivalent.*

1. *For every  $j \in \{1, \dots, n\}$ ,  $\beta p_j(\nu) = \alpha$ .*
2. *For every  $Q \in \alpha$ ,  $\bigcap_{j=1}^n p_j^{-1}(Q) \in \nu$ .*

Furthermore, these conditions hold for  $\nu$  with respect to some  $\alpha$  if, and only if, for every finite colouring  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  of  $\mathbb{N}^k$ , where  $\ell \geq 1$ , we have

$$\bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) \in \nu.$$

Also, for any two finite colourings  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  and  $\mathcal{Q}' = (Q'_1, \dots, Q'_{\ell'})$  of  $\mathbb{N}^k$ , where  $\ell, \ell' \geq 1$ , there exists a finite colouring  $\mathcal{Q}'' = (Q''_1, \dots, Q''_{\ell''})$  of  $\mathbb{N}^k$ , where  $\ell'' \geq 1$  such that

$$\bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) \cap \bigcup_{i=1}^{\ell'} \bigcap_{j=1}^n p_j^{-1}(Q'_i) = \bigcup_{i=1}^{\ell''} \bigcap_{j=1}^n p_j^{-1}(Q''_i).$$

*Proof.* For (1) implies (2), let us assume that, for every  $j \in \{1, \dots, n\}$ ,  $\beta p_j(\nu) = \alpha$ . Then, for every  $j \in \{1, \dots, n\}$ , and every  $Q \in \alpha$ , we have  $p_j^{-1}(Q) \in \nu$ , which implies that  $\bigcap_{j=1}^n p_j^{-1}(Q) \in \nu$ .

For (2) implies (1), fixing  $j \in \{1, \dots, n\}$ , we prove that  $\alpha \subseteq \beta p_j(\nu)$ , which is enough to prove that they are equal since ultrafilters are maximal for inclusion. Fix  $Q \in \alpha$ . By (2) we have that  $\bigcap_{k=1}^n p_k^{-1}(Q) \in \nu$ , and since  $\bigcap_{k=1}^n p_k^{-1}(Q) \subseteq p_j^{-1}(Q)$ , we deduce by the fact that  $\nu$  is an up-set that  $p_j^{-1}(Q) \in \nu$ , and thus that  $Q \in \beta p_j(\nu)$ .

We now treat the last assertion. On the one hand assume that there exists  $\alpha \in \beta(\mathbb{N}^k)$  satisfying (2). Since  $\alpha$  is an ultrafilter, for every finite colouring  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  of  $\mathbb{N}^k$ , where  $\ell \geq 1$ , there exists  $k \in \{1, \dots, \ell\}$  such that  $Q_k \in \alpha$ . Therefore,  $\bigcap_{j=1}^n p_j^{-1}(Q_k) \in \nu$ , and since

$$\bigcap_{j=1}^n p_j^{-1}(Q_k) \subseteq \bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i),$$

we deduce by up-set that this last subset belongs to  $\nu$ .

On the other hand, assume that for every finite colouring  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  of  $\mathbb{N}^k$ , where  $\ell \geq 1$ ,  $\bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i)$  is in  $\nu$ . We need to prove the existence of an ultrafilter  $\alpha \in \beta(\mathbb{N}^k)$  such that (2) holds. We set

$$\alpha := \{Q \subseteq \mathbb{N}^k : \bigcap_{j=1}^n p_j^{-1}(Q) \in \nu\}.$$

We prove that  $\alpha$  is an ultrafilter. First, since  $\nu$  is an ultrafilter, it does not contain the empty set, and thus  $\alpha$  does not contain the empty set. Also, since inverse image preserves finite intersections and inclusion, we deduce that  $\alpha$  is indeed a filter on  $\mathbb{N}^k$ . Furthermore, for any  $Q \subseteq \mathbb{N}^k$ ,  $(Q, Q^c)$  is a finite colouring of  $\mathbb{N}^k$ , thus, by (3),

$$\bigcap_{j=1}^n p_j^{-1}(Q) \cup \bigcap_{j=1}^n p_j^{-1}(Q^c) \in \nu.$$

This union being disjoint, and  $\nu$  being an ultrafilter, we deduce that exactly one element in  $\{\bigcap_{j=1}^n p_j^{-1}(Q), \bigcap_{j=1}^n p_j^{-1}(Q^c)\}$  belongs to  $\nu$ , in other words exactly one element in  $\{Q, Q^c\}$  is in  $\alpha$ .

Finally, we prove that

$$\{\bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) : (Q_1, \dots, Q_\ell) \text{ is a finite colouring of } \mathbb{N}^k, \ell \geq 1\}$$

is closed under intersection. Fix  $(Q_1, \dots, Q_\ell)$  and  $(Q'_1, \dots, Q'_{\ell'})$  two finite colourings of  $\mathbb{N}$ , for some  $\ell, \ell' \geq 1$ . We consider a finite colouring which refines both  $\mathcal{Q}$  and  $\mathcal{Q}'$ :  $\mathcal{Q}'' = (Q''_{i,j})_{\substack{1 \leq i \leq \ell \\ 1 \leq j \leq \ell'}}$ , the finite colouring of  $\mathbb{N}$ , such that, for every  $(i, j) \in \{1, \dots, \ell\} \times \{1, \dots, \ell'\}$ ,

$$Q''_{i,j} := Q_i \cap Q'_j.$$

We obtain that

$$\begin{aligned}
\bigcup_{\substack{1 \leq i \leq \ell \\ 1 \leq j \leq \ell'}} \bigcap_{j=1}^n p_j^{-1}(Q''_{i,j}) &= \bigcup_{\substack{1 \leq i \leq \ell \\ 1 \leq j \leq \ell'}} \bigcap_{j=1}^n p_j^{-1}(Q_i \cap Q'_j) \\
&= \bigcup_{\substack{1 \leq i \leq \ell \\ 1 \leq j \leq \ell'}} \bigcap_{j=1}^n (p_j^{-1}(Q_i) \cap p_j^{-1}(Q'_j)) \\
&= \bigcup_{1 \leq i \leq \ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) \cap \bigcup_{1 \leq i \leq \ell'} \bigcap_{j=1}^n p_j^{-1}(Q'_i).
\end{aligned}$$

□

We now treat the other condition involved in the definition of the family of ultrafilter equations  $\mathcal{E}_{u=v}^{p_1, \dots, p_n}$  from Definition 3.4, that is

$$L \models (\beta u(\nu) \leftrightarrow \beta v(\nu)).$$

For any set  $S$ , and any two subsets  $T_1, T_2$  of  $S$  we denote by  $T_1 \Delta T_2$  their *symmetric difference*, that is the subset of  $S$  such that, for any  $s \in S$ ,  $s \notin T_1 \Delta T_2$  if, and only if the condition

$$s \in T_1 \iff s \in T_2$$

holds.

**Lemma 3.11.** *For any set  $S$ , any two subsets  $T_1, T_2 \subseteq S$  and any ultrafilter  $\gamma \in \beta(S)$ , the following conditions are equivalent.*

1.  $T_1 \in \gamma$  if, and only if,  $T_2 \in \gamma$ .
2.  $(T_1 \Delta T_2)^c \in \gamma$ .

*In particular, for any  $k, n \geq 1$ , consider an ultrafilter  $\nu \in \beta(A^* \otimes (\mathbb{N}^k)^n)$ , two maps  $u, v : A^* \otimes (\mathbb{N}^k)^n \rightarrow A^*$ , and a language  $L \subseteq A^*$ . Then,  $L$  satisfies the ultrafilter equation  $\beta u(\nu) \leftrightarrow \beta v(\nu)$ , if, and only if  $E_{L,u,v} \in \nu$ , where*

$$E_{L,u,v} := \{(w, \bar{i}_1, \dots, \bar{i}_n) \in A^* \otimes (\mathbb{N}^k)^n : u(w, \bar{i}_1, \dots, \bar{i}_n) \in L \iff v(w, \bar{i}_1, \dots, \bar{i}_n) \in L\}.$$

*Proof.* Let us assume that  $T_1 \in \gamma$  if, and only if,  $T_2 \in \gamma$ . Since  $\gamma$  is closed under finite intersections, this last statement is equivalent to saying that  $T_1 \cap T_2 \in \gamma$  or  $T_1^c \cap T_2^c \in \gamma$ , and thus, since  $\gamma$  is an ultrafilter, equivalent to  $(T_1 \cap T_2) \cup (T_1^c \cap T_2^c) \in \gamma$ . By definition of the symmetric difference,

$$T_1 \Delta T_2 = ((T_1 \cap T_2) \cup (T_1^c \cap T_2^c))^c,$$

thus this is equivalent to saying that  $(T_1 \Delta T_2)^c \in \gamma$ .

The final statement is a simple application of this result for  $S = A^* \otimes (\mathbb{N}^k)^n$ ,  $T_1 = u^{-1}(L)$

and  $T_2 = v^{-1}(L)$ . □

We are now ready to give a reformulation of the equations of Definition 3.4 in terms of a condition relative to finite colourings of  $\mathbb{N}^k$ .

**Proposition 3.12.** *For any  $k, n \geq 1$ , any maps  $p_1, \dots, p_n : A^* \otimes (\mathbb{N}^k)^n \rightarrow \mathbb{N}^k$  and  $u, v : A^* \otimes (\mathbb{N}^k)^n \rightarrow A^*$ , a language  $L \subseteq A^*$  satisfies  $\mathcal{E}_{u=v}^{p_1, \dots, p_n}$  if, and only if, there exists a finite colouring  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  of  $\mathbb{N}^k$  for some  $\ell \geq 1$  such that*

$$\bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) \subseteq E_{L,u,v}.$$

*Proof.* For any language  $L \subseteq A^*$ , to satisfy the ultrafilter equation  $\mathcal{E}_{u=v}^{p_1, \dots, p_n}$  amounts to the following condition.

$$\forall \nu \in \beta(A^* \otimes (\mathbb{N}^k)^n), [\beta p_1(\nu) = \dots = \beta p_n(\nu) \implies (L \models \beta u(\gamma) \leftrightarrow \beta v(\gamma))].$$

By applying Lemma 3.10 and Lemma 3.11, we can reformulate this condition as follows.

$$\forall \nu \in \beta(A^* \otimes (\mathbb{N}^k)^n), (\{\bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) : (Q_1, \dots, Q_\ell) \text{ is a colouring of } \mathbb{N}^k, \ell \geq 1\} \subseteq \nu \implies E_{L,u,v} \in \nu)$$

In particular, we know that a filter is the intersection of all of the ultrafilters which contain it. Therefore, since Lemma 3.10 allows us to consider the filter

$$\mathcal{F} := \uparrow \{ \bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) : (Q_1, \dots, Q_\ell) \text{ is a finite colouring of } \mathbb{N}^k, \ell \geq 1 \},$$

we can simplify our condition into  $E_{L,u,v} \in \mathcal{F}$ , that is, there exists a finite colouring  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  of  $\mathbb{N}^k$  for some  $\ell \geq 1$  such that

$$\bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) \subseteq E_{L,u,v}.$$

□

Let us assume that a language  $L$  satisfies two families of ultrafilter equations,  $\mathcal{E}_{u=v}^{p_1, \dots, p_n}$  and  $\mathcal{E}_{u'=v'}^{p'_1, \dots, p'_{n'}}$ . By applying Proposition 3.12, this is equivalent to assuming the existence of two finite colourings  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  and  $\mathcal{Q}' = (Q'_1, \dots, Q'_{\ell'})$ , for some  $\ell, \ell' \geq 1$ , such that

$$\bigcup_{i=1}^{\ell} \bigcap_{j=1}^n p_j^{-1}(Q_i) \subseteq E_{L,u,v}$$

and

$$\bigcup_{i=1}^{\ell'} \bigcap_{j=1}^{n'} (p'_j)^{-1}(Q'_i) \subseteq E_{L,u',v'}.$$

Now, considering a common refinement  $\mathcal{Q}'' = (Q''_1, \dots, Q''_{\ell''})$  of the two colourings  $\mathcal{Q}$  and  $\mathcal{Q}'$ , for some  $\ell'' \geq 1$ , we obtain in particular that

$$\bigcup_{i=1}^{\ell''} \bigcap_{j=1}^n p_j^{-1}(Q''_i) \subseteq E_{L,u,v}$$

and

$$\bigcup_{i=1}^{\ell''} \bigcap_{j=1}^n (p'_j)^{-1}(Q''_i) \subseteq E_{L,u',v'}.$$

This allows for a reformulation of the ultrafilter equations introduced in Definition 3.5, 3.6 and 3.7.

**Corollary 3.13.** *For any language  $L \subseteq A^*$ , the following statements hold.*

1.  *$L$  satisfies  $\mathcal{E}_{ab=ba}$  and  $\mathcal{E}_{aab=abb}$ , for every  $a, b \in A$ , if, and only if, there exists a finite colouring  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  of  $\mathbb{N}$ , with  $\ell \geq 1$ , such that, for every  $a, b \in A$ ,*

$$\begin{cases} \bigcup_{i=1}^{\ell} A^* \otimes Q_i^2 \subseteq E_{L,f_{a,b},f_{b,a}} \\ \bigcup_{i=1}^{\ell} A^* \otimes Q_i^3 \subseteq E_{L,f_{a,a,b},f_{a,b,b}} \end{cases}.$$

2.  *$L$  satisfies  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  and  $\mathcal{E}_{a=aa}$ , for every  $a, b \in A$ , if, and only if, there exists a finite colouring  $\mathcal{Q} = (Q_1, \dots, Q_\ell)$  of  $\mathbb{N}$ , with  $\ell \geq 1$ , such that, for every  $a, b \in A$ ,*

$$\begin{cases} \bigcup_{i=1}^{\ell} A^* \otimes Q_i^2 \subseteq E_{L,f_{a,b},f_{b,a}} \\ \bigcup_{i=1}^{\ell} A^* \otimes Q_i^3 \subseteq E_{L,f_{a,a,b},f_{a,b,b}} \\ \bigcup_{i=1}^{\ell} L_{Q_i} \otimes Q_i \subseteq E_{L,f_a,f_{a,a}} \end{cases},$$

where, for every  $i \in \{1, \dots, \ell\}$ ,

$$L_{Q_i} \otimes Q_i := \{(w, j) \in A^* \otimes \mathbb{N} : j \in Q_i \text{ and } |w| \in Q_i\}.$$

**Example 3.14.** As a first application of these reformulations in terms of finite colouring of  $\mathbb{N}$ , we check that the couples of ultrafilters in the family  $\mathcal{E}_{ab=ba}$ , for every  $a, b \in A$  are indeed in the kernel of the continuous quotient  $q : \beta(A^*) \twoheadrightarrow X_1$ , dual to the canonical embedding  $\mathcal{B}_1 \hookrightarrow \mathcal{P}(A^*)$ .

Recall that we proved in Proposition 2.7 that the map  $q$  sends any ultrafilter  $\gamma \in \beta(A^*)$

to the family of closed subsets  $(C_d(\gamma))_{d \in A}$ , where, for every  $d \in A$ ,

$$C_d(\gamma) := \bigcap_{L_{\square_Q^d} \in \gamma} \widehat{Q}.$$

Now, fix  $a, b \in A$ , and consider  $\nu \in \beta(A^* \otimes \mathbb{N}^2)$ , such that  $\beta\pi_1(\nu) = \beta\pi_2(\nu)$ : we prove that  $\beta f_{a,b}(\nu)$  and  $\beta f_{b,a}(\nu)$  have the same image under  $q$ . Fix  $d \in A$ . First, note that

$$\begin{aligned} L_{\square_Q^d} \in \beta f_{a,b}(\nu) &\iff f_{a,b}^{-1}(L_{\square_Q^d}) \in \nu \\ &\iff \{(w, j_1, j_2) \in A^* \otimes \mathbb{N}^2 : w((j_1, j_2) \rightarrow (a, b)) \in L_{\square_Q^d}\} \in \nu \\ &\iff \{(w, j_1, j_2) \in A^* \otimes \mathbb{N}^2 : c_d(w((j_1, j_2) \rightarrow (a, b))) \subseteq Q\} \in \nu. \end{aligned}$$

Now, since  $\beta\pi_1(\nu) = \beta\pi_2(\nu)$ , we use Lemma 3.10 with the colouring  $(Q, Q^c)$  of  $\mathbb{N}$ , and we obtain that

$$(A^* \otimes Q^2) \cup (A^* \otimes (Q^c)^2) \in \gamma.$$

Therefore, by intersection,

$$\{(w, j_1, j_2) \in A^* \otimes \mathbb{N}^2 : c_d(w((j_1, j_2) \rightarrow (a, b))) \subseteq Q\} \in \nu$$

is equivalent to

$$\begin{aligned} &\{(w, j_1, j_2) \in A^* \otimes \mathbb{N}^2 : j_1, j_2 \in Q \text{ and } c_d(w((j_1, j_2) \rightarrow (a, b))) \subseteq Q\} \cup \\ &\{(w, j_1, j_2) \in A^* \otimes \mathbb{N}^2 : j_1, j_2 \in Q^c \text{ and } c_d(w((j_1, j_2) \rightarrow (a, b))) \subseteq Q^c\} \in \nu. \end{aligned}$$

Finally, observe that, for any  $j_1, j_2 \in Q$ ,

$$c_d(w((j_1, j_2) \rightarrow (a, b))) = c_d(w((j_1, j_2) \rightarrow (b, a))),$$

and the same holds for any  $j_1, j_2 \in Q^c$ . This allows us to prove that, for any  $d \in A$ ,

$$L_{\square_Q^d} \in \beta f_{a,b}(\nu) \iff L_{\square_Q^d} \in \beta f_{b,a}(\nu),$$

and thus, for every  $d \in A$ ,  $C_d(\beta f_{a,b}(\nu)) = C_d(\beta f_{b,a}(\nu))$  which means that  $q(\beta f_{a,b}(\nu)) = q(\beta f_{b,a}(\nu))$ .

### 3.3 A topological proof for ultrafilter equations on $\mathcal{B}_{0,1}$ and $\mathcal{B}_1$

In this section, we explain how it is possible to use Corollary 3.13 in order to show that the ultrafilter equations we introduced in Definition 3.6, 3.6 and 3.7 allow for a description of  $\mathcal{B}_{0,1}$  and  $\mathcal{B}_1$ .

### 3.3.1 Soundness

We start by proving that the Boolean algebras  $\mathcal{B}_{0,1}$  and  $\mathcal{B}_1$  satisfy the family of equations we introduced in Section 3.2. The reformulation of these equations we gave in Corollary 3.13, combined with the knowledge of the family of generators that we introduced in Proposition 3.2 and Corollary 3.3 for the Boolean algebras enables us to check it in a straight-forward fashion.

**Proposition 3.15.** *Let  $L \subseteq A^*$  be a language, then:*

1. *if  $L$  is in  $\mathcal{B}_{0,1}$ , then  $L$  satisfies the ultrafilter equations in the families  $\mathcal{E}_{ab=ba}$  and  $\mathcal{E}_{aab=abb}$ , for every  $a, b \in A$ ;*
2. *if  $L$  is in  $\mathcal{B}_1$ , then  $L$  satisfies the ultrafilter equations in the families  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  and  $\mathcal{E}_{a=aa}$ , for every  $a, b \in A$ .*

The idea behind this proposition is the following. All of the ultrafilter equations we introduced can be reformulated in terms of a condition that holds for a certain finite colouring of  $\mathbb{N}$ . This condition involves some finite words, and some of their positions. We check that the properties hold for finite words, and that this allows for the property to transfer at the level of ultrafilters.

*Proof.* (1): By Corollary 3.3, we know that  $\mathcal{B}_{0,1}$  is the Boolean algebra generated by the languages  $L_P$  and  $K_{Q,\bar{B}}$  where  $P \subseteq \mathbb{N}$ ,  $Q$  is a  $\ell$  colours colouring of  $\mathbb{N}$  for some  $\ell \geq 1$ , and  $\bar{B} \in \mathcal{P}(A)^\ell$ . Therefore, it is enough to prove that these languages satisfy the ultrafilter equations in question. We do so by using the reformulation of these equations introduced in Corollary 3.13. Fix  $P \subseteq \mathbb{N}$ , and  $a, b \in A$ , and consider the two colours colouring  $(P, P^c)$ . Fix  $(w, j_1, j_2) \in A^* \otimes P^2$ . We have that

$$\begin{aligned} w((j_1, j_2) \rightarrow (a, b)) \in L_P &\iff |w((j_1, j_2) \rightarrow (a, b))| \in P \\ &\iff |w((j_1, j_2) \rightarrow (b, a))| \in P \\ &\iff w((j_1, j_2) \rightarrow (b, a)) \in L_P. \end{aligned}$$

Since the same reasoning holds for any  $(w, j_1, j_2) \in A^* \otimes (P^c)^2$ , we have that

$$(A^* \otimes P^2) \cup (A^* \otimes (P^c)^2) \subseteq E_{L, f_{a,b}, f_{b,a}},$$

and we conclude by Corollary 3.13 that  $L_P$  satisfies the ultrafilter equation  $\mathcal{E}_{ab=ba}$ . We prove, in the exact same way, that  $L_P$  satisfies the ultrafilter equation  $\mathcal{E}_{aab=abb}$ : the argument holds since, for any  $(w, j_1, j_2, j_3) \in A^* \otimes \mathbb{N}^3$ ,

$$|f_{a,a,b}(w, j_1, j_2, j_3)| = |w| = |f_{a,b,b}(w, j_1, j_2, j_3)|.$$

Now, fix a finite colouring  $Q$  of  $\mathbb{N}$  with  $\ell$  colours, where  $\ell \geq 1$ ,  $\bar{B} \in \mathcal{P}(A)^\ell$  and  $a, b \in A$ .



Let us start by proving that

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^2 \subseteq E_{K_{\mathcal{Q}, \bar{B}}, f_{a,b}, f_{b,a}}.$$

Fix  $i \in \{1, \dots, \ell\}$ . We prove that every  $(w, j_1, j_2) \in A^* \otimes Q_i^2$  belongs to  $E_{K_{\mathcal{Q}, \bar{B}}, f_{a,b}, f_{b,a}}$ . On the one hand, notice that, for every  $i' \in \{1, \dots, \ell\}$ , with  $i \neq i'$ , the content of  $w((j_1, j_2) \rightarrow (a, b))$  on  $Q_{i'}$  is the same than the content of  $w((j_1, j_2) \rightarrow (b, a))$  on  $Q_{i'}$ : indeed, they are both equal to the content of  $w$  on  $Q_{i'}$ . On the other hand, the content of  $w((j_1, j_2) \rightarrow (a, b))$  on  $Q_i$  is the same than the content of  $w((j_1, j_2) \rightarrow (b, a))$  on  $Q_i$ : we only switched the letters at position  $j_1$  and  $j_2$ , which does not add nor remove any letter out of the content on  $Q_i$ . We deduce that, for every  $i \in \{1, \dots, \ell\}$ ,  $A^* \otimes Q_i^2 \subseteq E_{K_{\mathcal{Q}, \bar{B}}, f_{a,b}, f_{b,a}}$ , thus

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^2 \subseteq E_{K_{\mathcal{Q}, \bar{B}}, f_{a,b}, f_{b,a}},$$

and we conclude by Corollary 3.13 that  $K_{\mathcal{Q}, \bar{B}}$  satisfies the ultrafilter equation  $\mathcal{E}_{ab=ba}$ .

The argument is almost exactly the same to prove that  $K_{\mathcal{Q}, \bar{B}}$  satisfies the ultrafilter equation  $\mathcal{E}_{aab=abb}$ . Fix  $i \in \{1, \dots, \ell\}$ . We prove that every  $(w, j_1, j_2, j_3) \in A^* \otimes Q_i^3$  belongs to  $E_{K_{\mathcal{Q}, \bar{B}}, f_{a,a,b}, f_{a,b,b}}$ . On the one hand, notice that, for every  $i' \in \{1, \dots, \ell\}$ , with  $i \neq i'$ , the content of  $w((j_1, j_2, j_3) \rightarrow (a, a, b))$  on  $Q_{i'}$  is the same than the content of  $w((j_1, j_2, j_3) \rightarrow (a, b, b))$  on  $Q_{i'}$ : indeed, they are both equal to the content of  $w$  on  $Q_{i'}$ . On the other hand, the content of  $w((j_1, j_2, j_3) \rightarrow (a, a, b))$  on  $Q_i$  is the same than the content of  $w((j_1, j_2, j_3) \rightarrow (a, b, b))$  on  $Q_i$ : we only replaced the occurrence of  $a$  available at the position  $j_2 \in Q_i$  by an occurrence of  $b$ , which was already available at the position  $j_3 \in Q_i$ . The letter  $a$  is still available at the position  $j_1 \in Q_i$ : the content on  $Q_i$  has not been altered. We deduce that

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^3 \subseteq E_{K_{\mathcal{Q}, \bar{B}}, f_{a,a,b}, f_{a,b,b}},$$

and we conclude by Corollary 3.13 that  $K_{\mathcal{Q}, \bar{B}}$  satisfies the ultrafilter equation  $\mathcal{E}_{aab=abb}$ .

(2): By Corollary 3.3, we know that  $\mathcal{B}_1$  is the Boolean algebra generated by the languages  $K_{\mathcal{Q}, \bar{B}}$ , where  $\mathcal{Q}$  is a  $\ell$  colours colouring of  $\mathbb{N}$  for some  $\ell \geq 1$ , and  $B \in \mathcal{P}(A)^\ell$ . By (1), we only need to prove that the languages of this form satisfy the ultrafilter equations  $\mathcal{E}_{a=a.a}$ , for every  $a \in A$ . By Corollary 3.13, we do so by proving that

$$\bigcup_{i=1}^{\ell} L_{Q_i} \otimes Q_i \subseteq E_{K_{\mathcal{Q}, \bar{B}}, f_{a,a}.a}.$$

Fix  $i \in \{1, \dots, \ell\}$ . We prove that every  $(w, j) \in A^* \otimes Q_i$  such that  $|w| \in Q_i$  belongs to  $E_{K_{\mathcal{Q}, \bar{B}}, f_{a,a}.a}$ . On the one hand, notice that, for every  $i' \in \{1, \dots, \ell\}$ , with  $i \neq i'$ , the content of  $w(j \rightarrow a)$  on  $Q_{i'}$  is the same than the content of  $w(j \rightarrow a).a$  on  $Q_{i'}$ , since  $|w|$  does

not belong to  $Q_i$ . On the other hand, the content of  $w(j \rightarrow a)$  on  $Q_i$  is the same than the content of  $w(j \rightarrow a).a$  on  $Q_i$ : the only difference between these two words is the presence of one more occurrence of  $a$  on a position which belongs to  $Q_i$ , but since  $j \in Q_i$ , the letter  $a$  is already present in the content of both of these words on  $Q_i$ . We deduce that

$$\bigcup_{i=1}^{\ell} L_{Q_i} \otimes Q_i \subseteq E_{K_{\mathcal{Q}, \bar{B}}, f_a, f_a.a},$$

and we conclude that  $K_{\mathcal{Q}, \bar{B}}$  satisfies the ultrafilter equation  $\mathcal{E}_{a=a.a}$ .  $\square$

### 3.3.2 Completeness for $\mathcal{B}_1$

We prove that the languages satisfying the family of ultrafilter equations  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  and  $\mathcal{E}_{a=a.a}$ , for every  $a, b \in A$ , all belong to  $\mathcal{B}_1$ . In order to do so, we use the presentation of  $\mathcal{B}_1$  we provided in Proposition 3.2. As explained in Section 3.1, the languages generating  $\mathcal{B}_1$  are of the form  $K_{\mathcal{Q}, \bar{B}}$ , where  $\mathcal{Q}$  is a finite colouring of  $\mathbb{N}$  with  $\ell$  colours, for some  $\ell \geq 1$ , and  $\bar{B} \in \mathcal{P}(A)^\ell$ . For a fixed finite colouring  $\mathcal{Q}$ , the finitely indexed equivalence relation  $\sim_{\mathcal{Q}}$  introduced in Section 3.1 gives a finite partition of  $A^*$ . We prove that, if a language  $L \subseteq A^*$  satisfies all of our ultrafilter equations, then it is possible to find a finite colouring  $\mathcal{Q}$  such that  $L$  can be written as a finite union of some of the equivalence classes for  $\sim_{\mathcal{Q}}$ .

We start by detailing the situation when the colouring in question only has two colours in the following Lemma. We then generalize the argument to any finite colouring of  $\mathbb{N}$  in Proposition 3.17.

**Lemma 3.16.** *Let  $L \subseteq A^*$  be a language satisfying  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  and  $\mathcal{E}_{a=a.a}$ , for every  $a, b \in A$ . Let us consider a finite colouring  $\mathcal{Q}$  of  $\mathbb{N}$  satisfying the condition from Corollary 3.13 (2), and assume that  $\mathcal{Q} = (Q, Q^c)$  for some  $Q \subseteq \mathbb{N}$ . Then, for any words  $w, w' \in A^*$  such that  $w \sim_{\mathcal{Q}} w'$  and, for every  $i \in Q^c \cap \{0, \dots, \min(|w|, |w'|) - 1\}$ ,  $w_i = w'_i$ , we have that*

$$w \in L \iff w' \in L.$$

Since the proof is quite technical, we give an overview of our proof method. First, we have to provide some intuition about how we can use the ultrafilter equations we required in Lemma 3.16. We do so for the ultrafilter equations  $(\mathcal{E}_{ab=ba})_{a,b \in A}$ . Let us assume that a language  $L$  satisfies  $\mathcal{E}_{ab=ba}$ , for every  $a, b \in A$ . Let us consider a finite colouring  $\mathcal{Q}$  of  $\mathbb{N}$  satisfying the condition from Corollary 3.13 (2), and assume that  $\mathcal{Q} = (Q, Q^c)$  for some  $Q \subseteq \mathbb{N}$ . Fix two integers  $i, j \in Q$  and a word  $w$  such that  $|w| > \max(i, j)$ . Since  $L$  satisfies, in particular,  $\mathcal{E}_{w_i w_j = w_j w_i}$ , Corollary 3.13 (2) implies that  $(w, i, j) \in E_{L, f_{w_i, w_j}, f_{w_j, w_i}}$ , that is,  $w$  is in  $L$  if, and only if, the word obtained out of  $w$  by exchanging  $w_i$  and  $w_j$  is in  $L$ . We can keep on applying this reasoning to construct a word  $w'$  that is in  $L$  if, and only if  $w$  is in  $L$  and such that:

- $w'$  and  $w$  have the same length;

- for every  $i \in Q^c$ ,  $w_i = w'_i$ ;
- for every  $a \in A$ , and every  $i \in Q$ ,  $|w|_a = |w'|_a$ .

This is the main idea we use. In Lemma 3.16, in order to prove that  $w$  is in  $L$  if, and only if,  $w'$  is in  $L$ , we will construct a family of intermediate words  $v_1, \dots, v_n$ , for some  $n \in \mathbb{N}$ . Setting  $v_0 := w$  and  $v_{n+1} := w'$ , we will prove that, for every  $i \in \{0, \dots, n\}$ ,  $v_i \in L$  if, and only if,  $v_{i+1} \in L$ . These intermediate words will be constructed out of  $w$ , using the ultrafilter equations we introduced, and will all have the same content than  $w$  on  $Q$  and on  $Q^c$ .

*Proof.* We consider a language  $L \subseteq A^*$  such that there exists  $Q \subseteq \mathbb{N}$  such that, for every  $a, b \in A$ ,

$$(A^* \otimes Q^2) \cup (A^* \otimes (Q^c)^2) \subseteq E_{L, f_{a,b}, f_{b,a}},$$

$$(A^* \otimes Q^3) \cup (A^* \otimes (Q^c)^3) \subseteq E_{L, f_{a,a,b}, f_{a,b,b}},$$

and

$$(L_Q \otimes Q) \cup (L_{Q^c} \otimes Q^c) \subseteq E_{L, f_{a,a}}.$$

We consider two finite words  $w, w' \in A^*$  such that  $w \sim_Q w'$  and for every  $i \in Q^c \cap \{0, \dots, \min(|w|, |w'|) - 1\}$ ,  $w_i = w'_i$ . We set  $N := \min(|w|, |w'|)$ . We want to prove that  $w$  is in  $L$  if, and only if,  $w'$  is in  $L$ . In order to do so, we start by defining the following families of endofunctions of  $A^*$ .

$$\mathcal{M}_{\mathcal{E}, 1, Q}^{ab=ba} := \{(f_{a,b}^{j_1, j_2}, f_{b,a}^{j_1, j_2}) : j_1, j_2 \in Q \cap \{0, \dots, N-1\}, a, b \in A\},$$

where, for every  $j_1, j_2 \in Q \cap \{0, \dots, N-1\}$ , and every  $a, b \in A$ ,  $f_{a,b}^{j_1, j_2} : A^* \rightarrow A^*$  sends any finite word  $v$  to  $f_{a,b}(v, j_1, j_2)$  if  $(v, j_1, j_2)$  is in  $A^* \otimes \mathbb{N}^2$ , and to  $v$  otherwise. We define in an analogous way

$$\mathcal{M}_{\mathcal{E}, 1, Q}^{aab=abb} := \{(f_{a,a,b}^{j_1, j_2, j_3}, f_{a,b,b}^{j_1, j_2, j_3}) : j_1, j_2, j_3 \in Q \cap \{0, \dots, N-1\}, a, b \in A\},$$

$$\mathcal{M}_{\mathcal{E}, 1, Q}^{a=a.a} := \{(f_a^{j_1}, f_a^{j_1}.a) : j_1 \in Q \cap \{0, \dots, N-1\}, a \in A\},$$

and

$$\mathcal{M}_{\mathcal{E}, 1, Q} := \mathcal{M}_{\mathcal{E}, 1, Q}^{ab=ba} \cup \mathcal{M}_{\mathcal{E}, 1, Q}^{aab=abb} \cup \mathcal{M}_{\mathcal{E}, 1, Q}^{a=a.a}.$$

Now, let us assume that there exists  $n \in \mathbb{N}$ , that there exists a family of pairs of maps  $(f_m, g_m)_{1 \leq m \leq n}$  and a finally, that there exists a finite sequence of words  $(v_m)_{1 \leq m \leq n}$ , satisfying the following properties:

$$\left. \begin{array}{l} 1. v_1 = w \text{ and } v_n = w'; \\ 2. m \in \{1, \dots, n\}, (f_m, g_m) \in \mathcal{M}_{\mathcal{E},1,Q} \\ 3. \text{ for every } m \in \{1, \dots, n\}, f_m(v_m) = v_m; \\ 4. \text{ for every } m \in \{1, \dots, n-1\}, g_m(v_m) = v_{m+1}, \text{ and } \\ \quad g_n(v_n) = v_n. \end{array} \right\} (*)$$

In particular, if we denote composition of maps by  $\circ$ , we would have that

$$w' = \left( \bigcirc_{m=n}^1 g_m \right)(w),$$

and therefore by Corollary 3.13 (1), we are able to conclude that  $w$  is in  $L$  if, and only if  $w'$  is in  $L$ . The end of this proof describes how it is possible to construct  $n, (f_m, g_m)_{1 \leq m \leq n}$  and  $(v_m)_{1 \leq m \leq n}$ , satisfying  $(*)$ .

We start by treating the case where  $w$  and  $w'$  have the same length, that is  $N$ . In that case, there exists  $n_1, \dots, n_p < N$  for some  $p \in \mathbb{N}$  such that

$$Q \cap \{0, \dots, N-1\} = \{n_1, \dots, n_p\},$$

and we consider the finite words  $w|_Q := w_{n_1} \dots w_{n_p}$  and  $w'|_Q := w'_{n_1} \dots w'_{n_p}$ . We assumed that  $w \sim_Q w'$ , thus we set  $B := \langle w, Q \rangle = \langle w', Q \rangle$ . Notice that  $w|_Q$  and  $w'|_Q$  are equal if, and only if,  $w$  and  $w'$  are equal.

First, assume that each letter in  $B$  occurs exactly once in  $w|_Q$  and  $w'|_Q$ . In that case, these two words only differ by the order of their letters, and so do  $w$  and  $w'$ . If  $w|_Q$  and  $w'|_Q$  are the same word, then we are done. Otherwise, there are at least two distinct letters  $a, b \in B$  which occur in  $w|_Q$  and  $w'|_Q$ , and yet do not occupy the same position in both words. We thus consider the unique integers  $s_1, s_2$  and  $t_1, t_2$  in  $\{n_1, \dots, n_p\}$  such that  $w_{n_{s_1}} = w_{n_{t_1}} = a$  and  $w'_{n_{s_2}} = w'_{n_{t_2}} = b$ , and we know that  $(s_1, s_2) \neq (t_1, t_2)$ . Then,  $w = f_{a,b}(w, n_{s_1}, n_{t_1})$  and the word  $w_1 = f_{b,a}(w, n_{s_1}, n_{t_1})$  is such that the positions of the occurrences of  $a$  in  $w_1$  coincide with the one of  $w'$ . Since any permutation on a finite set can be written as a finite product of transpositions, we can iterate this reasoning finitely many times, and eventually construct a finite family  $(f_m, g_m)_{1 \leq m \leq n} \subseteq \mathcal{M}_{\mathcal{E},1,Q}^{ab=ba}$  and a finite sequence of words  $(v_m)_{1 \leq m \leq n}$  for some  $n \in \mathbb{N}$  satisfying  $(*)$ .

Now, assume there exists a letter which occurs strictly more than once in  $w|_Q$  or  $w'|_Q$ , say  $w|_Q$  without any loss of generality. If, for every  $a \in B$ ,  $|w|_Q|_a = |w'|_Q|_a$ , then these two words only differ by the order of appearance of each of their letters. By applying a similar reasoning to the case where  $w|_Q$  and  $w'|_Q$  have exactly one occurrence of each letter in  $A$ , we can construct a finite family  $(f_m, g_m)_{1 \leq m \leq n} \subseteq \mathcal{M}_{\mathcal{E},1,Q}^{ab=ba}$  and a finite sequence of words  $(v_m)_{1 \leq m \leq n}$ , for some  $n \in \mathbb{N}$  satisfying the four conditions we mentioned. Else, there exists a letter  $a \in B$  such that  $|w|_Q|_a \neq |w'|_Q|_a$ , say  $|w|_Q|_a > |w'|_Q|_a$  without any loss of generality. Pick  $n_{s_1}, n_{s_2} \in \{n_1, \dots, n_p\}$  distincts such that  $w_{j_1} = w_{j_2} = a$ . Now, since

$w|_Q$  and  $w'|_Q$  have the same length, there exists a letter  $b \in B$  with  $|w|_Q|_b < |w'|_Q|_b$ , and  $b \neq a$ . Of course, since  $w \sim_Q w'$ , there exists  $n_{s_3} \in \{n_1, \dots, n_p\}$  such that  $w_{j_3} = b$ . Then,  $w = f_{a,a,b}(w, n_{s_1}, n_{s_2}, n_{s_3})$  and the word  $w_1 := f_{a,b,b}(w, n_{s_1}, n_{s_2}, n_{s_3})$  satisfies the following properties:

- $w \sim_Q w_1$ ;
- for every  $i \in Q^c \cap \{0, \dots, N-1\}$ ,  $(w_1)_i = w_i$ ;
- by Corollary 3.13 (2),  $w$  is in  $L$  if, and only if,  $w_1$  is in  $L$ ;
- $|w_1|_a = |w|_a - 1$ .

We can keep on using this argument, until we obtain a word which brings us back to the case where  $|w|_Q|_a = |w'|_Q|_a$ . By applying this argument to every letter of the alphabet  $A$ , we construct a finite word  $w'' \in A^*$  such that  $w'' \sim_Q w'$ , for every  $i \leq N$ ,  $w''_i = w'_i$  and, for every  $a \in B$ ,  $|w''|_a = |w'|_Q|_a$ : this situation has already been treated previously.

Finally, let us treat the case where  $w$  and  $w'$  do not have the same length.

Without any loss of generality, let us assume that  $|w| < |w'|$ . We are going to prove the following statement: for any  $n \in \{|w|, \dots, |w'|\}$ , there exists a finite word  $v \in A^*$  satisfying the following conditions:

- $|v| = n$ ;
- $w \in L$  if, and only if,  $v \in L$ ;
- $v \sim_Q w$  and, for every  $i \in Q^c \cap \{0, \dots, n-1\}$ ,  $w'_i = v_i$ .

Proving this is enough for us to conclude. Indeed, this will mean, in particular, that there exists a word  $w''$ , with same length than  $w'$ , such that, for every  $i \in Q^c \cap \{0, \dots, |w'|\}$ ,  $w'_i = w''_i = w_i$ ,  $w'' \sim_Q w \sim_Q w'$ , and  $w'' \in L$  if, and only if,  $w \in L$ , thus if, and only if,  $w' \in L$ . We will then be able to use the reasoning conducted at the very beginning of the proof to  $w''$  and  $w'$ , as they have the same length.

Let us now prove the statement. For the base case, that is, when  $n = |w|$ , we can use the word  $w$ , as it is trivial that it satisfies the condition we require. For the induction step, let us assume that there exists a finite word  $v$  with length  $n \in \{|w|, \dots, |w'|\}$  satisfying the conditions we require. We will use this word to construct a word which also satisfies these conditions, and has exactly one more letter than  $v$ .

Since  $(Q, Q^c)$  is a colouring of  $\mathbb{N}$ , we make a case distinction depending on which color  $|v|$  belongs to. If  $Q \cap \{0, \dots, |v| - 1\}$  is non-empty, pick an integer  $m$  in this subset, and set

$$a := \begin{cases} v_m & \text{if } |v| \in Q \\ w'_{|v|} & \text{if } |v| \in Q^c \text{ or } Q \cap \{0, \dots, |v| - 1\} = \emptyset \end{cases}.$$

and

$$j := \begin{cases} m & \text{if } |v| \in Q \\ |v| & \text{if } |v| \in Q^c \text{ or } Q \cap \{0, \dots, |v| - 1\} = \emptyset \end{cases}.$$

Since  $(v, j)$  belongs to  $(L_Q \otimes Q) \cup (L_{Q^c} \otimes Q^c)$ , we have by Corollary 3.13 (3) that  $v$  is in  $L$  if, and only if,  $v.a$  is in  $L$ . Observe that, by construction,  $w \sim_Q v \sim_Q v.a$ , and for every  $i \in Q^c \cap \{0, \dots, n\}$ ,  $w'_i = (v.a)_i$ . This allows us to conclude the induction.  $\square$

We now generalize the argument to any finite colouring of  $\mathbb{N}$ .

**Proposition 3.17.** *Let  $L \subseteq A^*$  be a language satisfying  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  and  $\mathcal{E}_{a=aa}$ , for every  $a, b \in A$ . Let us consider a finite colouring  $\mathcal{Q}$  of  $\mathbb{N}$  satisfying the condition from Corollary 3.13(2). Then, for any words  $w, w' \in A^*$  such that  $w \sim_{\mathcal{Q}} w'$ , we have*

$$w \in L \iff w' \in L.$$

*Proof.* We consider a language satisfying this property, that is,  $L \subseteq A^*$  such that there exists a finite colouring  $\mathcal{Q}$  of  $\mathbb{N}$  with  $\ell$  colours, where  $\ell \geq 1$ , such that, for every  $a, b \in A$ ,

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^2 \subseteq E_{L, f_{a,b}, f_{b,a}},$$

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^3 \subseteq E_{L, f_{a,a,b}, f_{a,b,b}},$$

and

$$\bigcup_{i=1}^{\ell} L_{Q_i} \otimes Q_i \subseteq E_{L, f_a, f_{a \cdot a}},$$

by Corollary 3.13 (2). We consider two finite words  $w, w' \in A^*$  such that  $w \sim_{\mathcal{Q}} w'$ . We want to prove that  $w$  is in  $L$  if, and only if,  $w'$  is in  $L$ .

We can apply the reasoning we used in Proof 3.3.2 independently to each of the subsets  $Q_1, \dots, Q_{\ell}$ . More precisely, for every  $i \in \{1, \dots, \ell\}$ , we consider  $\mathcal{M}_{\mathcal{E}, 1, Q_i}$  as introduced in Proof 3.3.2. We can construct  $\ell$  independent integers  $(n_i)_{1 \leq i \leq \ell}$ ,  $\ell$  independent families of pairs of maps  $((f_m^i, g_m^i)_{1 \leq m \leq n_i})_{1 \leq i \leq \ell}$  and  $\ell$  independent finite sequence of words  $((v_m^i)_{1 \leq m \leq n_i})_{1 \leq i \leq \ell}$  satisfying the following properties: for every  $i \in \{1, \dots, \ell\}$ ,

1.  $(v_1^i)_{|Q_i} = w_{|Q_i}$  and  $(v_{n_i}^i)_{Q_i} = w'_{Q_i}$ ;
2. for every  $m \in \{1, \dots, n_i\}$ ,  $(f_m^i, g_m^i) \in \mathcal{M}_{\mathcal{E}, 1, Q_i}$ ;
3. for every  $m \in \{1, \dots, n_i\}$ ,  $f_m^i(v_m^i) = v_m^i$ ;
4. for every  $m \in \{1, \dots, n_i - 1\}$ ,  $g_m^i(v_m^i) = v_{m+1}^i$ , and  $g_{n_i}^i(v_{n_i}^i) = v_{n_i}^i$ .

In particular, if we denote composition of maps by  $\circ$ ,

$$w' = \left( \bigcirc_{i=1}^{\ell} \bigcirc_{m=n_i}^1 g_m^i \right)(w),$$

and thus, by Lemma 3.16,  $w$  belongs to  $L$  if, and only if,  $w'$  belongs to  $L$ .  $\square$

We are now ready to prove that every language  $L \subseteq A^*$  satisfying our equations is necessarily such that there exists a finite colouring  $\mathcal{Q}$  of  $\mathbb{N}$  such that  $L$  saturates the finitely indexed equivalence relation  $\sim_{\mathcal{Q}}$ , and thus  $L$  can be expressed as a Boolean combinations of the generators of  $\mathcal{B}_1$ .

**Corollary 3.18.** *Any language  $L \subseteq A^*$  which satisfies the ultrafilter equations  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  and  $\mathcal{E}_{a=a.a}$ , for every  $a, b \in A$ , is in  $\mathcal{B}_1$ .*

*Proof.* Consider a language  $L \subseteq A^*$  satisfying the equations  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  and  $\mathcal{E}_{a=a.a}$ , for every  $a, b \in A$ . By Proposition 3.12, there exists a finite colouring  $(Q_1, \dots, Q_\ell)$  of  $\mathbb{N}$ , for some  $\ell \geq 1$ , such that, for every  $a, b \in A$ ,

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^2 \subseteq E_{L, f_{a,b}, f_{b,a}},$$

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^3 \subseteq E_{L, f_{a,a,b}, f_{a,b,b}}$$

and

$$\bigcup_{i=1}^{\ell} L_{Q_i} \otimes Q_i \subseteq E_{L, f_a, f_{a.a}}.$$

Fix a word  $w \in L$ , then there exists a unique  $\bar{B} \in \mathcal{P}(A)^\ell$  such that  $w \in K_{\mathcal{Q}, \bar{B}}$ . By Proposition 3.17, for any  $w' \in K_{\mathcal{Q}, \bar{B}}$ , we have that  $w' \in L$ . More generally, we deduce that the finitely indexed equivalence relation  $\sim_{\mathcal{Q}}$  on  $A^*$ , is saturated by  $L$ . Therefore, there exists a finite family  $\mathfrak{B} \subseteq \mathcal{P}(A)^\ell$  such that

$$L = \bigcup_{\bar{B} \in \mathfrak{B}} K_{\mathcal{Q}, \bar{B}},$$

and by Proposition 3.2, we conclude that  $L$  is in  $\mathcal{B}_1$ .  $\square$

### 3.3.3 Completeness for $\mathcal{B}_{0,1}$

We prove that the languages satisfying the family of ultrafilter equations  $\mathcal{E}_{ab=ba}$ ,  $\mathcal{E}_{aab=abb}$  for every  $a, b \in A$ , all belong to  $\mathcal{B}_{0,1}$ . The idea is similar to the reasoning we provided to prove Corollary 3.18, the difference being that the equivalence relation that we introduced in the previous subsection is now insufficient. Indeed, since we added the family of languages  $L_P$ , for any  $P \subseteq \mathbb{N}$ , we need to take into account the length of the words we consider. This has consequences on the family of ultrafilter equations we want to look at. For instance, for any  $a \in A$ , consider a language  $L$  which satisfies the family of equations  $\mathcal{E}_{a=a.a}$ . This corresponds, informally, to the idea that, if a word  $w$  belongs to  $L$ , then it is possible to construct a longer word by adding a certain letter to  $w$ , which will still belong to  $L$ . In this sense, the length of a word of  $L$  is not relevant, and

thus we expect that this family of equations will not be required in order to describe  $\mathcal{B}_{0,1}$ . We begin by providing an analogous for Proposition 3.17 which takes this observation into account.

**Proposition 3.19.** *Let  $L \subseteq A^*$  be a language satisfying  $\mathcal{E}_{ab=ba}$  and  $\mathcal{E}_{aab=abb}$ , for every  $a, b \in A$ . Let us consider a finite colouring  $\mathcal{Q}$  of  $\mathbb{N}$  satisfying the condition from Corollary 3.13 (1). Then, for any words  $w, w' \in A^*$  of same length such that  $w \sim_{\mathcal{Q}} w'$ , we have*

$$w \in L \iff w' \in L.$$

*Proof.* The proof uses the exact same method that was used in Proof 3.3.2 and Proof 3.3.2, except that we always assume that the length of both words involved is the same.  $\square$

We are now ready to prove that every language  $L \subseteq A^*$  satisfying  $\mathcal{E}_{ab=ba}$  and  $\mathcal{E}_{aab=abb}$ , for every  $a, b \in A$  is necessarily such that there exists a finite colouring  $\mathcal{Q}$  of  $\mathbb{N}$  such that  $L$  saturates the finitely indexed equivalence relation  $\sim_{\mathcal{Q}}$ , and thus  $L$  can be expressed as a Boolean combinations of the generators of  $\mathcal{B}_{0,1}$ .

**Corollary 3.20.** *Any language  $L \subseteq A^*$  which satisfies the ultrafilter equations  $\mathcal{E}_{ab=ba}$  and  $\mathcal{E}_{aab=abb}$ , for every  $a, b \in A$ , is in  $\mathcal{B}_{0,1}$ .*

*Proof.* Consider a language  $L \subseteq A^*$  satisfying the equations  $\mathcal{E}_{ab=ba}$  and  $\mathcal{E}_{aab=abb}$ , for every  $a, b \in A$ . By Proposition 3.12, there exists a finite colouring  $(Q_1, \dots, Q_\ell)$  of  $\mathbb{N}$ , for some  $\ell \geq 1$ , such that, for every  $a, b \in A$ ,

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^2 \subseteq E_{L, f_{a,b}, f_{b,a}}$$

and

$$\bigcup_{i=1}^{\ell} A^* \otimes Q_i^3 \subseteq E_{L, f_{a,a,b}, f_{a,b,b}}.$$

Fix a word  $w \in L$ . There exists a unique  $\bar{B} \in \mathcal{P}(A)^\ell$  such that  $w \in K_{\mathcal{Q}, \bar{B}}$ . By Proposition 3.19, for any  $w' \in K_{\mathcal{Q}, \bar{B}}$ , such that  $|w| = |w'|$  we have that  $w' \in L$ . More generally, there exists a finite family  $\mathfrak{B} \subseteq \mathcal{P}(A^\ell)$  such that, if we set

$$P := \{n \in \mathbb{N} : L \cap A^n \neq \emptyset\},$$

then

$$L = \left( \bigcup_{\bar{B} \in \mathfrak{B}} K_{\mathcal{Q}, \bar{B}} \right) \cap \left( \bigcup_{n \in P} A^n \right).$$

Finally, since

$$\bigcup_{n \in P} A^n = L_P,$$



we have that

$$L = \left( \bigcup_{\bar{B} \in \mathfrak{B}} K_{Q, \bar{B}} \right) \cap L_P,$$

and by Proposition 3.2, we conclude that  $L$  is in  $\mathcal{B}_{0,1}$ . □

## Conclusion of the chapter

In this chapter, we provided a new method to prove the completeness and the soundness of the ultrafilter equations that were introduced in [29] for the fragment of first-order logic corresponding to sentences written by only using nullary and unary uniform numerical predicates. Our method being more general, it allowed us to also formulate more ultrafilter equations, and these allowed us to characterise the Boolean algebra of languages corresponding to sentences written by only using unary uniform numerical predicates.

A lot of the remarks we made in the conclusion of the previous chapter still apply here. Indeed, a first step in order to get an intuition over the ultrafilter equations for a Boolean algebra is to compute this algebra's dual space. No work currently available treats the topic of ultrafilter equations for the Boolean algebra  $\mathcal{B}\Sigma_1[\mathcal{N}_k^u]$ , for a fixed  $k \in \mathbb{N}$ . This should be the next natural step following this work. We believe that the approach that consists in reformulating the equations in terms of a condition that involves considering every finite colouring of  $\mathbb{N}^k$ , for some  $k \in \mathbb{N}$  could be a good approach in order to generalize Corollary 3.18.

- [1] J. Almeida. *Finite Semigroups and Universal Algebra*. Series in algebra. World Scientific, 1994.
- [2] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, USA, 2009.
- [3] S. Awodey. *Category Theory*. Oxford University Press, Inc., USA, 2nd edition, 2010.
- [4] D. Barrington. Non-uniform automata over groups. *Information and Computation*, 89(2):109–132, 1990.
- [5] N. Bezhanishvili, L. Carai, and P. Morandi. Coalgebras for the powerset functor and thomason duality. *CoRR*, abs/2008.01849, 2020.
- [6] G. Bezhanishvili and A. Kurz. Free modal algebras: A coalgebraic perspective. In *Algebra and Coalgebra in Computer Science*, pages 143–157, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.
- [7] N. Bezhanishvili and M. Gehrke. Finitely generated free heyting algebras via birkhoff duality and coalgebra. *Log. Methods Comput. Sci.*, 7(2), 2011.
- [8] P. Blackburn. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2001.
- [9] M. Bojanczyk. Recognisable languages over monads. In *Proceedings of the 19th International Conference on Developments in Language Theory*, volume 9168 of *Lecture Notes in Computer Science*, pages 1–13. Springer International Publishing, 2015.
- [10] M.M. Bonsangue. Duality beyond sober spaces: Topological spaces and observation frames. *Theoretical Computer Science*, 151(1):79–124, 1995. Selected Papers of the Workshop on Topology and Completion in Semantics.

- [11] C. Borlido, S. Czarnetzki, M. Gehrke, and A. Krebs. Stone duality and the substitution principle. In *26th EACSL Annual Conference on Computer Science Logic, CSL 2017, August 20-24, 2017, Stockholm, Sweden*, volume 82 of *LIPICs*, pages 13:1–13:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.
- [12] M. Branco. Varieties of languages. In *Semigroups, Algorithms, Automata and Languages*, pages 91–132. World Scientific, 2002.
- [13] S. Burris and H.P Sankappanavar. *A Course in Universal Algebra*. Springer, 1981.
- [14] C.C Chang. Algebraic analysis of many valued logics. *Transactions of the American Mathematical Society*, 88(2):467–490, 1958.
- [15] L. Chaubard, J.-E. Pin, and H. Straubing. First order formulas with modular predicates. pages 211–220. IEEE, 2006.
- [16] T. Colcombet and D. Petrisan. Automata in the Category of Glued Vector Spaces. In Kim G. Larsen, Hans L. Bodlaender, and Jean-Francois Raskin, editors, *42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017)*, volume 83 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 52:1–52:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.
- [17] T. Colcombet and D. Petrisan. Automata minimization: a functorial approach. *Log. Methods Comput. Sci.*, 16(1), 2020.
- [18] S. Czarnetzki and A. Krebs. Using duality in circuit complexity, 2015.
- [19] R.J Daverman and R.B Sher. *Handbook of Geometric Topology*. Elsevier Science & Technology Books, 2002.
- [20] A. Davey and H. A. Priestley. *Introduction to lattices and order*. Cambridge University Press, Cambridge, 1990.
- [21] M. Dekkers. Stone duality: An application in the theory of formal languages. *Master’s thesis*, 2008.
- [22] V. Diekert and M. Kufleitner. Fragments of first-order logic over infinite words, 2009.
- [23] S. Eilenberg. Automata, languages, and machines. In *Pure and applied mathematics*, 1974.
- [24] L. Esakia. Topological kripke models. *Soviet Math.* 15, pages 147–151, 1974.
- [25] R. Fagin. Generalized first-order spectra, and polynomial. time recognizable sets. *SIAM-AMS Proc.*, 7, 01 1974.

- [26] J. Font. *Abstract Algebraic Logic - An Introductory Textbook*. College Publications, 2016.
- [27] M. Gehrke, S. Grigorieff, and J.-E. Pin. Duality and equational theory of regular languages. In *ICALP*, 2008.
- [28] M. Gehrke and A. Krebs. Stone duality for languages and complexity. *SIGLOG News*, 4(2):29–53, 2017.
- [29] M. Gehrke, A. Krebs, and J.-E. Pin. From ultrafilters on words to the expressive power of a fragment of logic. *Lecture Notes in Computer Science*, 8614, 2014.
- [30] M. Gehrke, D. Petrisan, and L. Reggio. The schützenberger product for syntactic spaces. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55, pages 112:1–112:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016.
- [31] M. Gehrke, D. Petrisan, and L. Reggio. Quantifiers on languages and codensity monads. In *32nd Annual Symposium on Logic in Computer Science*, 2017.
- [32] S. Givant. *Duality Theories for Boolean Algebras with Operators*. Springer Monographs in Mathematics. Springer International Publishing, 2014.
- [33] S. Givant and P.R Halmos. *Introduction to Boolean Algebras*. Undergraduate Texts in Mathematics. Springer New York, 2008.
- [34] P. Harpe. Topics in geometric group theory. *Bibliovault OAI Repository, the University of Chicago Press*, 01 2000.
- [35] C. Hartonas. Duality for modal  $\mu$ -logics. *Theoretical Computer Science*, 202:193–222, 07 1998.
- [36] N. Immerman. Descriptive complexity. *Graduate Texts in Computer Science*, 1999.
- [37] J. Isbell. Atomless parts of spaces. *Mathematica Scandinavica*, 31:5–32, Jun. 1972.
- [38] P. Johnstone. *Stone Spaces*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1986.
- [39] D. Kozen, K. G. Larsen, R. Mardare, and P. Panangaden. Stone duality for markov processes. In *In Proceedings of the 28th Annual IEEE Symposium on Logic in Computer Science: LICS*.
- [40] M. Kunc. Equational description of pseudovarieties of homomorphisms. *RAIRO - Theoretical Informatics and Applications*, 37(3):243–254, 2003.

- [41] C. Kupke, A. Kurz, and V. Venema. Stone coalgebras. *Theor. Comput. Sci.*, 327(1-2):109–134, 2004.
- [42] M. Lawson. Finite automata. In *Handbook of Networked and Embedded Control Systems*, pages 117–144. Birkhäuser, 2005.
- [43] L. Libkin. Elements of finite model theory (texts in theoretical computer science. an eatcs series), 2004.
- [44] J. B. Saxe M. Furst and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. In *Proceedings of the 22nd Annual Symposium on Foundations of Computer Science*, page 260–270, USA, 1981. IEEE Computer Society.
- [45] A. Maciel, P. Péladéau, and D. Thérien. Programs over semigroups of dot-depth one. *Theoretical Computer Science*, 245(1):135–148, 2000.
- [46] V. Marra and L. Spada. The dual adjunction between mv-algebras and tychonoff spaces. *Studia Logica: An International Journal for Symbolic Logic*, 100(1/2):253–278, 2012.
- [47] E. Michael. Topologies on spaces of subsets. *Transactions of the American Mathematical Society*, 71, 1951.
- [48] A. Nerode. Linear automaton transformations. *Proceedings of the American Mathematical Society*, 9(4):541–544, 1958.
- [49] C. Papadimitriou. *Computational complexity*. Addison-Wesley, 1994.
- [50] S. Perifel. *Complexité algorithmique*. Ellipses, 2014.
- [51] J.-E. Pin. *Syntactic Semigroups*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.
- [52] J.-E. Pin. Expressive power of existential first-order sentences of büchi’s sequential calculus. 291:155–174, 2005.
- [53] J.-E. Pin. Equational descriptions of languages. *Int. J. Found. Comput. Sci.*, 23:1227–1240, 2011.
- [54] J.-E. Pin. Mathematical foundations of automata theory, 2012.
- [55] J.-E. Pin and H. Straubing. Some results on C-varieties. *RAIRO - Theoretical Informatics and Applications (RAIRO: ITA)*, 39:239–262, 2005.
- [56] J.-E. Pin and P. Weil. Profinite semigroups, Mal’cev products and identities. *Journal of Algebra*, 182(3):604–626, 1996.
- [57] J.-E. Pin and P. Weil. A reiterman theorem for pseudovarieties of finite first-order structures, 1996.

- [58] N. Pippenger. Regular languages and stone duality. *Theory of Computing Systems*, 30(2):121–134, 1997.
- [59] T. Place and M. Zeitoun. The tale of the quantifier alternation hierarchy of first-order logic over words. *SIGLOG News*, 2:4–17, 2015.
- [60] L.S. Pontryagin. The theory of topological commutative groups. *Annals of Mathematics Second Series*, Vol. 35, No. 2 (Apr., 1934), pp. 361–388, pages 361–388, 1934.
- [61] J. Reiterman. The birkhoff theorem for finite algebras. *algebra universalis*, 14:1–10, 1982.
- [62] J. Rutten. Universal coalgebra: a theory of systems. *Theoretical Computer Science*, 249(1):3–80, 2000. Modern Algebra.
- [63] J. Sakarovitch. *Elements of Automata Theory*. Cambridge University Press, USA, 2009.
- [64] M.P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8(2):190–194, 1965.
- [65] S. Grigorieff, M. Gehrke, and J.-E. Pin. A topological approach to recognition. In *ICALP*, 2010.
- [66] I. Simon. Piecewise testable events. In *Proceedings of the 2nd GI Conference on Automata Theory and Formal Languages*, page 214–222, Berlin, Heidelberg, 1975. Springer-Verlag.
- [67] M. Stone. The theory of representation for boolean algebras. *Transactions of the American Mathematical Society*, 40(1):37–111, 1936.
- [68] M. Stone. Topological representations of distributive lattices and brouwerian logics. *Journal of Symbolic Logic*, 3(2):90–91, 1938.
- [69] H. Straubing. Constant-depth periodic circuits. *International Journal of Algebra and Computation*, 01(01):49–87, 1991.
- [70] H. Straubing. *Finite Automata, Formal Logic, and Circuit Complexity*. Birkhauser Verlag, Basel, Switzerland, Switzerland, 1994.
- [71] H. Straubing and P. Weil. An introduction to finite automata and their connection to logic. pages 3–43, 2012.
- [72] P. Tesson and D. Thérien. Logic meets algebra: the case of regular languages. *Logical Methods in Computer Science*, 3(1), Feb 2007.

- [73] D. Thérien and T. Wilke. Nesting until and since in linear temporal logic. In Helmut Alt and Afonso Ferreira, editors, *STACS 2002*, pages 455–464, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.
- [74] A. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, s2-42(1):230–265, 01 1937.
- [75] M. Urquhart. Duality for algebras of relevant logics. *Studia Logica: An International Journal for Symbolic Logic*, 56(1/2):263–276, 1996.
- [76] L. Vietoris. Bereiche zweiter ordnung. *Monatshefte für Mathematik und Physik*, 32:258–280.
- [77] H. Vollmer. *Introduction to Circuit Complexity: A Uniform Approach*. Springer-Verlag, Berlin, Heidelberg, 1999.
- [78] P. Walker. *The Stone-Cech compactification*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge. Springer-Verlag Berlin Heidelberg, 1974.
- [79] P. Weil. Pro methods in semigroup theory. *Int. J. Algebra Comput*, 12:137–178, 2000.
- [80] R. Williams. Nonuniform acc circuit lower bounds. *J. ACM*, 61(1), January 2014.