

Dossier de candidature

Au poste de Maître de Conférences en Informatique

2025-03-25

Résumé : Ce document présente ma candidature pour le poste de Maître de Conférences en Informatique à l'Institut Polytechnique de Bordeaux. Il se décompose en quatre sections indépendantes : mon parcours académique et mes expériences professionnelles, mes activités d'enseignement, mes travaux de recherche et mon projet de recherche et d'intégration. Chaque section est brièvement résumée dans les encadrés ci-dessous et détaillée dans les sections suivantes. Une table des matières est disponible en page 2.

POSTE

Offre 251816
Section 27
Laboratoire LaBRI
Équipe LX (Méthodes et Modèles Formels)
Établissement Institut Polytechnique de Bordeaux
Département Informatique

CANDIDAT

Nom Aliaume LOPEZ
Mél ad.lopez@uw.edu.pl
Tél +33 6 94 42 07 71
Adresse Teofila Lenartowicza 16
01-614, Varsovie, Pologne
Site <https://www.irif.fr/~alopez>



PARCOURS

2023 – 2025 **Postdoctorat**
Université de Varsovie
avec Mikołaj Bojańczyk
2020 – 2021 **Développeur Fullstack**
Autorité de Sécurité Nucléaire
En césure de thèse
2019 – 2023 **Doctorat**
IRIF et LMF
Lauréat de 2 prix internationaux
2018 **Agrégation Mathématiques**
Option Info. Classé 5e
2015 – 2019 **Masters et Licences**
ENS Paris-Saclay
Stages de recherche
Birmingham, Ljubljana, Paris

ENSEIGNEMENT

2023 – 2024 **Chargé de TD (M1)**
Université de Varsovie
Transducteurs
2022 – 2023 **Chargé de TP (L1)**
Université Paris Cité
Python, Java, Systèmes
2021 – 2022 **Chargé de TD/TP (L2, L3)**
Université Paris Cité
OCaml, Projet, Réseaux
2019 – 2020 **Chargé de TD (L3, M1)**
ENS Paris-Saclay
Complexité, Programmation
2017 – 2019 **Khôlleur de Mathématiques**
Lycée Lakanal, Inst. St. Marie
MPSI / MP

RECHERCHE

8 **Publications (conférences)**
LICS, CSL, STACS, FoSSaCS, ...
2 **Publications (journaux)**
3 **Préprints**
PC Member CSL 2026

PROJET ET INTÉGRATION

Axe 1 **Systèmes de Transitions**
Graphes, Beaux Préordres, Bases de données
Axe 2 **Programmes à états finis**
Automates, Logique, Transductions
Intégration : LX, RATIO

☐ Ce document utilise **knowledge** : une notion pointée vers sa *définition*.

Table des matières

1	Parcours	3
1.1	Formation	3
1.2	Expérience professionnelle	3
2	Enseignement	5
2.1	Enseignement universitaire	5
2.2	Intégration au sein de l'équipe pédagogique	6
3	Recherche	7
3.1	Thèmes de recherche	7
3.2	Beaux préordres et espaces Noëthériens	8
3.2.1	De nouveaux espaces	9
3.2.2	L'ensemble des parties finies	10
3.3	Théorie des modèles finis	10
3.3.1	Espaces préspectraux logiquement présentés	10
3.3.2	Localité	11
3.4	Théorie des automates et transductions	12
3.4.1	Fonctions $\mathbb{Z}$ -polyrégulières	13
3.4.2	Fonctions $\mathbb{N}$ -polyrégulières	13
3.5	Travaux en cours	14
3.6	Rayonnement	14
4	Projet de Recherche & Intégration	16
4.1	Introduction	16
4.1.1	Direction 1 : Systèmes centrés sur les données	16
4.1.2	Direction 2 : Vérification de transformations	17
4.2	Direction 1 : systèmes centrés sur les données	19
4.3	Direction 2 : transductions	22
4.4	Intégration	25
4.4.1	Collaborations nationales et internationales	25
A	Attestation de réussite	33
B	Procès-verbal de soutenance	36
C	Rapport de thèse 1	43
D	Rapport de thèse 2	46
E	Rapport de soutenance	50
F	Ackermann Award 2024	52

1 Parcours

J'ai effectué un parcours universitaire à l'interface entre les mathématiques et l'informatique, domaines dans lesquels j'ai obtenu une licence et un master. Ma volonté de rester en contact avec ces deux matières explique la durée de mon master d'informatique (2016 — 2019), effectué en deux ans, avec une année de pause pour passer l'agrégation de mathématiques en même temps que le master de formation à l'enseignement supérieur en mathématiques (durant l'année scolaire 2017 — 2018). De même, la durée de mon doctorat (2019 — 2023) s'explique par une année de césure durant laquelle j'ai été développeur pour une administration publique française. Je suis actuellement en post-doctorat à l'Université de Varsovie (Pologne) où je travaille avec Mikołaj BOJAŃCZYK.

1.1 Formation

2023 — présent. POST-DOCTORAT EN INFORMATIQUE. (Sous la direction de Mikołaj Bojańczyk)
Université de Varsovie

2019 — 2023. DOCTORAT EN INFORMATIQUE. (Thèse soutenue le 12/09/2023). `hal`

« First Order Preservation Theorems in Finite Model Theory : Locality, Topology and Limit Constructions ».

Préparée sous la direction de :

— Jean Goubault-Larrecq (*PREX, Laboratoire Méthodes Formelles, École Normale Supérieure Paris-Saclay*).

— Sylvain Schmitz (*PR, Institut de Recherche en Informatique Fondamentale, Université Paris-Cité*).

Mon jury était composé de :

— Serenella Cerrito (*Professeure des Universités, Université Paris Saclay, Univ EVRY, France*).

— Anuj Dawar (*Professor, University of Cambridge*).

— Arnaud Durand (*Professeur des Universités, Université Paris-Cité*).

— Patrice Ossona de Mendez (*Chargé de Recherche, EHESS*).

— Daniela Petrisan (*Maitresse de Conférences, Université Paris-Cité*).

— Luca Reggio (*Senior Fellow Researcher, University College London*).

2018. AGRÉGATION DE MATHÉMATIQUES. Option Informatique (Classé 5ème sur 320 admis)
École Normale Supérieure de Cachan

2017 — 2018. MASTER DE MATHÉMATIQUES. Master de Formation à l'Enseignement Supérieur (Mention très bien)
École Normale Supérieure de Cachan

2016 — 2019. MASTER D'INFORMATIQUE. Master Parisien de Recherche en Informatique (Mention très bien)
École Normale Supérieure de Cachan

2015 — 2016. LICENCE D'INFORMATIQUE. (Mention très bien)
École Normale Supérieure de Cachan

2015 — 2016. LICENCE DE MATHÉMATIQUES. (Mention très bien)
École Normale Supérieure de Cachan

2013 — 2015. MPSI/MP*. Classe préparatoire aux grandes écoles (Mention très bien)
Lycée Joffre, Montpellier

2013. BACCALAURÉAT SCIENTIFIQUE. (Mention très bien)
Lycée Georges Pompidou, Castelnau-le-Lez

1.2 Expérience professionnelle

Mes expériences professionnelles relèvent majoritairement de l'enseignement supérieur et de la recherche, à l'exception d'une année effectuée en césure de thèse durant laquelle j'ai travaillé pour une administration publique en tant que développeur fullstack sur un projet de transformation numérique. J'ai démontré ma capacité à enseigner (aussi bien en tant que chargé de TD durant mon doctorat et postdoctorat, qu'en donnant des khôlles de mathématiques en MP et MPSI), m'intégrer dans des équipes de recherche internationales (Birmingham, Ljubljana, Varsovie), et produire des outils applicatifs concrets dans un court laps de temps (pour l'Autorité de Sûreté Nucléaire, via un portage de la Direction Interministérielle du Numérique).

Vous trouverez ci-après la liste détaillée de mes expériences professionnelles.

2023 – 2025. CONTRAT POST-DOCTORAL. CDD à temps plein. Sous la direction de Mikołaj Bojańczyk.

Université de Varsovie

Recherche théorique sur des langages de programmation minimalistes manipulant des chaînes de caractères. L'objectif est de comprendre leur expressivité, de fournir des méthodes d'optimisation pour ces programmes et de faire le lien avec d'autres modèles de calculs.

2019 – 2023. DOCTORANT EN INFORMATIQUE AVEC CHARGE D'ENSEIGNEMENT. CDD à temps plein. Sous la direction de Jean Goubault-Larrecq et Sylvain Schmitz.

École Normale Supérieure Paris-Saclay, Université Paris Cité

Recherche sur les systèmes centrés sur les données d'un point de vue très théorique.

2020 – 2021. CHARGÉ D'AFFAIRES - DÉVELOPPEUR FULLSTACK. CDD à temps plein.

Direction Interministérielle du Numérique, Autorité de Sûreté Nucléaire

Création d'une plateforme de collecte et d'analyse de documents publics relatifs à la sûreté nucléaire et à la radioprotection dans le cadre du programme "Entrepreneur d'Intérêt Général"

2017 – 2019. KHÔLLEUR DE MATHÉMATIQUES. Vacataire à temps partiel.

Institution Sainte Marie, Lycée Lakanal

Mars 2017 – Juillet 2017. STAGE DE RECHERCHE. Stage non rémunéré à temps plein. Sous la direction d'Alex K. Simpson.

University of Ljubljana, Slovénie

Création d'une méthode d'analyse de programmes prenant en compte les "effets de bords", qui sont essentiels dans les langages réels.

Juin 2015 – Juillet 2015. STAGE DE RECHERCHE. Stage non rémunéré à temps plein. Sous la direction de Dan R. Ghica.

University of Birmingham, Angleterre

Optimisation de circuits contenant des boucles de rétro action.

2015 – 2019. NORMALIEN ÉLÈVE. Fonctionnaire Stagiaire.

ENS Paris-Saclay

2 Enseignement

Durant mon doctorat, j'ai été chargé de TD successivement à l'École Normale Supérieure Paris-Saclay pour des L3 et M1, puis chargé de TP à l'Université Paris-Cité, pour des L1, L2 et L3. Par la suite, j'ai été chargé de TD d'un programme de M1 (cours optionnel) à Varsovie au second semestre de l'année scolaire 2023–2024. Cette diversité d'expériences me permet d'être à l'aise avec les différents enseignements, leurs formats, et les pédagogies adaptées aux différentes années.

2.1 Enseignement universitaire

Comme indiqué dans la partie expérience professionnelle, j'ai également donné des khôlles de mathématiques en MP et MPSI durant ma scolarité à l'ENS Paris-Saclay, et j'ai été contacté pour être examinateur aux oraux d'informatique du concours d'entrée aux Mines de Paris pour l'année 2025, mais cette activité complémentaire n'a malheureusement pas été acceptée par l'Université de Varsovie. J'espère pouvoir, dans l'éventualité d'un retour en France, continuer à m'investir dans ce type d'activités.

Vous trouverez ci-après la liste détaillée de mes charges d'enseignement universitaire.

2023 – 2024. Université de Varsovie. Monitorat. 24 heures équivalent TD.

- **Transducteurs (nouveau cours).** M1 Étudiants en informatique. Entre dix et quinze étudiants.

Création de sujets de TP et de devoirs maison.

2022 – 2023. Université Paris Cité. Monitorat (CDSN). 60 heures équivalent TD.

- **TP de Programmation en Python.** L1 Étudiants en licence Mathématiques et informatiques appliquées aux sciences humaines et sociales. Environ une quinzaine d'étudiants.

Surveillance d'examen, création et correction de "questions de cours" hebdomadaires

- **TP de Programmation en Java.** L1 Étudiants en licence d'informatique. Plus d'une vingtaine d'étudiants.

Surveillance d'examens, participation à la relecture et mise à jour des sujets de TP

- **TP d'Introduction aux Systèmes d'Exploitation.** L1 Étudiants en licence d'informatique. Plus d'une vingtaine d'étudiants.

Surveillance d'examens, participation à la relecture des sujets de TP

2021 – 2022. Université Paris Cité. Monitorat (CDSN). 60 heures équivalent TD.

- **TP de Programmation Fonctionnelle.** L3 Étudiants en licence d'informatique. Une petite vingtaine d'étudiants.

Jury dans les soutenances du projet de fin de cours

- **TP/TD de Gestion de Projet (Projet).** L2 Étudiants en licence d'informatique. Une petite vingtaine d'étudiants.

Jury dans les soutenances du projet de fin de cours

- **TP de Programmation Réseau.** L3 Étudiants en licence d'informatique. Une petite vingtaine d'étudiants.

Jury dans les soutenances de projet, création d'un outil de "stress test" pour les programmes des étudiants.

2019 – 2020. École Normale Supérieure Paris-Saclay. Monitorat (CDSN). 48 heures équivalent TD.

- **TD de Complexité Avancée.** M1 Étudiants et élèves normaliens. Entre dix et quinze étudiants.

Reprise des supports de TD, avec ajout et modification d'exercices Les supports sont disponibles sur ma page web <https://irif.fr/~alopez/enseignement.html#complexité-avancée>

- **TD de Programmation 1.** L3 Étudiants et élèves normaliens. Entre vingt et trente étudiants.

Reprise des supports de TD, avec ajout et modification d'exercices Les supports sont disponibles sur ma page web <https://irif.fr/~alopez/enseignement.html#programmation-1>

En termes de pédagogie, quelque chose que j'ai beaucoup aimé en TP de programmation python (qui se fait en petits groupes et au niveau L1) est de faire des sessions de live coding pour la correction de certains exercices, avec des élèves qui « expliquent ce qu'ils ont fait ». J'ai trouvé que cela permet deux choses intéressantes : d'une part, cela permet aux élèves qui ont des difficultés avec la syntaxe, ou sont plutôt lents avec l'ordinateur de participer, de l'autre, cela force les élèves à exprimer clairement à un autre être humain leurs intentions (et se mettre d'accord sur ce que je dois coder). La principale difficulté est de faire participer toute la classe, ce qui demande beaucoup d'efforts et un petit groupe d'élèves.

Une autre pratique que j'aime beaucoup est le "quizz" (de type Kahoot par exemple) où les élèves se mettent en équipe pour répondre à des questions projetées au tableau. L'avantage de cette méthode est que même sur une classe de 30/40 élèves tout le monde peut participer. Bien entendu, il faut mettre uniquement des questions "piège", sans quoi les élèves perdent rapidement intérêt dans l'exercice, qui lui-même ne doit pas durer plus d'une quinzaine de minutes.

Par ailleurs, je trouve un grand plaisir à enseigner, d'une part, car cela me pousse à apprendre de nouvelles connaissances parfois très éloignées de mes sujets de recherche (réseau, systèmes), d'autre part, car j'apprécie la relation avec les étudiants et le fait de les voir progresser.

2.2 Intégration au sein de l'équipe pédagogique

Je suis convaincu de ma capacité à enseigner et encadrer des étudiants de l'ENSEIRB-MATMECA en filière informatique. J'ai déjà enseigné la calculabilité et la complexité à l'ENS Paris-Saclay (ex ENS Cachan), des cours de programmation (Python, Java, OCaml) à l'Université Paris-Cité, ainsi que des cours de génie logiciel et de réseaux. Par ailleurs, j'ai été développeur fullstack à l'Autorité de Sécurité Nucléaire, et j'ai une appétence pour les méthodes agiles et le DevOps. De plus, j'ai obtenu l'agrégation de mathématiques option informatique, ce qui me permet d'enseigner les parties plus théoriques du Syllabus (algorithmique, mathématiques discrètes, logique, complexité, calculabilité, etc.). Enfin, mes domaines de recherche en lien avec les méthodes formelles et la vérification de programmes me tournent naturellement vers le cours de spécialisation du semestre 9 de l'ENSEIRB-MATMECA, Algorithmes et méthodes formelles (AMF).

Bien qu'ayant moins de facilités avec des cours plus opérationnels comme la cybersécurité, la robotique ou le calcul haute performance, je suis convaincu de ma capacité à monter en compétences sur ces sujets en fonction des besoins de l'école.

Pour conclure, je souhaiterais souligner que mon profil d'enseignant-chercheur est beaucoup plus orienté pratique que mon profil de recherche. Par exemple, j'utilise beaucoup de logiciels libres, je programme sur mon temps libre, et je suis très à l'aise avec la ligne de commande. J'apprécie beaucoup comprendre comment fonctionnent les choses et les expliquer à d'autres, et je me retrouve dans le Syllabus de l'ENSEIRB-MATMECA aux semestres 5, 6 et 7 qui mêlent algorithmique, mathématiques et programmation (systèmes, réseaux).

3 Recherche

Mon parcours de recherche a commencé par un intérêt pour la sémantique des langages de programmation, avec deux stages (un court, en M1, un long, en M2) portant sur cette thématique. Ces deux stages, l'un à Birmingham, l'autre à Ljubljana ont mené à des publications dans des conférences internationales. Suite à ces expériences, je me suis ré-orienté dans le domaine de la théorie des modèles finis et la combinatoire, mes principaux objets d'étude étant les beaux préordres et la logique au premier ordre sur des structures finies. C'est dans ces domaines que j'ai effectué ma thèse de doctorat en co-supervision entre l'École Normale Supérieure Paris-Saclay, et l'Université Paris-Cité. Ma thèse a reçu deux prix internationaux :

2024 **Ackermann Award** [\[lien\]](#)

Le prix de thèse Ackermann est le prix de thèse exceptionnelle de l'EACSL pour la logique en informatique.

2024 **E.W. Beth Dissertation Prize** [\[lien\]](#)

Depuis 1998, l'Association for Logic, Language, and Information (FoLLI) décerne chaque année le Prix de thèse E.W. Beth aux thèses de doctorat exceptionnelles en logique, langage et information, avec le soutien financier de la Fondation E.W. Beth.

Plus récemment, j'ai pris intérêt pour la médiation scientifique et j'ai participé à la fête de la science, organisée par l'IRIF en 2023. Dans une moindre mesure, mon activité de développeur à l'ASN est une expérience de transfert, même si mes recherches sur les modèles finis et la logique au premier ordre n'ont pas crucialement participé à la réalisation du produit. Cette expérience m'a beaucoup aidé à comprendre mes propres sujets de recherche d'un point de vue plus pratique.

Depuis mon arrivée à Varsovie fin 2023, je travaille sur des problèmes en théorie des automates, en particulier sur le transfert de propriétés de décidabilité des *langages réguliers* aux *fonctions* et l'expressivité des différents modèles de fonctions dans la littérature.

3.1 Thèmes de recherche

Résumé. Cette section synthétise les travaux de recherche effectués durant ma thèse de doctorat en co-supervision entre l'IRIF (Sylvain SCHMITZ) et le LMF (Jean GOUBAULT-LARRECQ), ainsi que ceux effectués durant ma première année de post-doctorat à l'Université de Varsovie sous la direction de Mikołaj BOJAŃCZYK. Ces travaux se situent à l'intersection de plusieurs domaines : la logique (au premier ordre et monadique du second ordre), la théorie de l'ordre (beaux préordres) et la topologie (espaces Noëthériens). Ces domaines de l'informatique fondamentale trouvent leurs applications au travers de la vérification de programmes et servent plus généralement de socle théorique aux approches par méthodes formelles.

Dans une première partie, je présente mes travaux sur les beaux préordres et les espaces Noëthériens, deux objets provenant respectivement de la théorie de l'ordre et de la topologie, qui permettent tous deux de garantir une forme de terminaison. En plus de leur intérêt théorique, l'utilisation des beaux préordres pour la vérification de systèmes de transitions possédant un nombre infini d'états, sous la forme des *Well-Structured Transition Systems* (WSTS), a été plébiscitée par deux prix de la communauté : le prix du « LICS Test of Time Award 2016 » pour le travail de ABDULLA et al. [\[Abd+96\]](#) sur les WSTS, et le prix « CAV 2017 » pour des contributions fondamentales à la vérification formelle assistée par ordinateur [\[Abd+17\]](#). Dans ce contexte, les espaces Noëthériens forment une généralisation des beaux préordres à un cadre topologique, qui permet l'adaptation de ces techniques à des ensembles pour lesquels la notion d'ordre est très peu adaptée : l'exemple prototype étant $\mathbb{C}^k$ muni de la topologie de Zariski, qui permet de parler de systèmes manipulant des *nombres*. Dans ce domaine, mes contributions sont multiples. D'une part, j'ai travaillé à l'obtention de bornes sur la *complexité* de certains beaux préordres et espaces Noëthériens, ce qui permet de mieux comprendre leurs propriétés mais aussi de *borner* le temps de calcul d'algorithmes utilisant ces structures. D'autre part, j'ai travaillé à l'obtention de *nouveaux* espaces Noëthériens (resp. beaux préordres) ce qui a permis d'étendre la portée de ces techniques de vérification.

Dans une seconde partie, je présente mes travaux sur la théorie des modèles finis. Ces travaux ont eu pour objectif initial d'adapter les techniques basées sur les beaux préordres et les espaces Noëthériens à des systèmes infinis dont l'espace des états est une classe de structures relationnelles finies (modélisant une base de données). L'exemple classique étant un serveur web possédant une base de donnée et qui fait évoluer celle-ci au cours de son exécution. Dans ce cadre, il est très rare que l'espace des états soit muni d'un beau préordre ou d'une topologie Noëthérienne permettant sa vérification. En revanche, les *transitions* de tels systèmes sont, en un certain sens, compatibles avec la logique du premier ordre (qui est la logique sous-jacente au langage SQL). De plus, les *propriétés à vérifier* sur ces systèmes sont également exprimables en logique du premier ordre. Ainsi, j'ai travaillé à

l'élaboration d'une théorie à l'intersection entre *logique* (propriétés exprimables en logique du premier ordre) et *topologie* (garanties de terminaison), ce qui a donné naissance aux espaces préspectraux logiquement présentés, un élément central de ma thèse de doctorat. Cette théorie est intimement liée aux théorèmes de préservation en logique du premier ordre, un domaine où j'ai significativement contribué par mon approche originale.

Dans une troisième partie, je regroupe mes travaux sur la théorie des automates et des transductions. Ces travaux sont, à l'origine, issus d'une collaboration spontanée avec Gaëtan DOUÉNEAU-TABOT, avec qui j'avais déjà travaillé sur des sujets non académiques (l'insertion des docteurs au sein de l'administration publique Française). Nous nous sommes intéressés à des questions de *décidabilité* sur des transductions *quantitatives* : en partant d'une requête $\varphi(x, y)$, on peut construire une fonction qui, à un modèle M associe le nombre de couples (a, b) dans M vérifiant $\varphi(x, y)$. Cette construction, qui n'est autre que le `SELECT COUNT(*) ... FROM ... WHERE` du langage SQL, est capturée dans le cas des mots finis par différents modèles de calculs dont l'interaction s'est avérée fructueuse (automates pondérés, MSO-quantitatif, fonctions polyrégulières). Ainsi, j'ai introduit avec Gaëtan DOUÉNEAU-TABOT et Thomas COLCOMBET la notion de *fonction $\mathbb{Z}$ -polyrégulières*, et continué leur étude durant mon post-doctorat. Ces travaux ont été l'occasion de découvrir un nouveau champ de recherche et de m'ouvrir à de nouvelles collaborations, notamment en faisant un post-doctorat sous la direction de Mikołaj BOJAŃCZYK. J'ai eu le plaisir d'y encadrer deux stagiaires et de mettre en place les séances de travaux dirigés d'un nouveau cours sur les transducteurs.

Enfin, dans une quatrième et dernière partie, j'expose mes travaux en cours. Ces travaux tirent profit de l'ensemble des compétences acquises durant ma thèse et mon post-doctorat, notamment en combinant des techniques acquises à l'Université de Varsovie pour résoudre certaines questions laissées en suspens durant ma thèse, mais aussi en posant de nouvelles questions qui nécessitent précisément cet ensemble de compétences, et qui semblent mener à des résultats significatifs dans le domaine des beaux préordres et de la vérification formelle.

Certains de mes travaux ne s'insèrent pas directement dans les dynamiques ci-dessus, par conséquent je ne les détaille pas. Ils sont listés dans l'index bibliographique.

3.2 Beaux préordres et espaces Noëthériens

Cette partie est principalement basée sur une publication à FoSSaCS 2023 [Lop23a] et deux publications dans des journaux [Abr+23; GHL23]. Ils ont fait l'objet de deux présentations en sus des conférences : l'une à la réunion du projet ANR BraVAS en avril 2022 et l'autre à un séminaire sur les beaux préordres à Lyon en février 2023.

Beaux préordres. Un *beau préordre* est un ensemble $(X, \leq)$ où $\leq$ est une relation binaire réflexive et transitive, telle que pour toute suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de X , il existe un couple d'indices $i < j$ telle que $x_i \leq x_j$. En tant qu'objet combinatoire, les beaux préordres apparaissent régulièrement [Kru72] dans des domaines variés de l'informatique, allant de la théorie des graphes avec le théorème de ROBERTSON et SEYMOUR [RS04], à la théorie des nombres [Hig52]. Néanmoins, ce qui a motivé ma recherche est l'application des beaux préordres aux preuves de terminaison d'algorithmes, et plus particulièrement à la vérification des "Systèmes de Transitions Bien Structurés" (WSTS), qui sont des systèmes de transition sur un ensemble infini d'états équipé d'un beau préordre, et pour lesquels la relation de transition est *compatible* avec le préordre sous-jacent [voir par exemple Abd+96; AJ98; FSo1].

Lors de la vérification de systèmes de transition, il est souvent utile de considérer l'ensemble des parties des états [Jan99; SF17]. Malheureusement, l'ensemble $\mathcal{P}(X)$ des parties d'un beau préordre X n'est pas toujours un beau préordre [Rad54]. Ce problème a été contourné de trois manières très différentes dans la littérature :

1. Une première solution *ad hoc* est de soigneusement éviter l'utilisation de $\mathcal{P}(X)$ et le remplacer, par exemple, par l'ensemble des parties finies $\mathcal{P}_f(X)$.
2. Une deuxième solution est de travailler avec des beaux préordres qui ne sont pas pathologiques.¹ C'est ce qui a motivé l'introduction des pré-meilleurs-ordres [Pou72; Mil85].
3. Une troisième et dernière approche a été proposée par GOUBAULT-LARRECQ [Gou07], qui a remarqué que la notion mathématique d'espace topologique Noëthérien pouvait jouer le rôle d'une généralisation des beaux préordres à un cadre topologique, dans lequel la construction de l'ensemble des parties n'est pas problématique.

1. Intuitivement, tels que $\mathcal{P}(X)$, $\mathcal{P}^2(X)$, ... sont tous des beaux préordres.

Espaces Noëthériens. Un espace topologique est un ensemble X muni d'une topologie $\tau \subseteq \mathcal{P}(X)$, c'est-à-dire d'une famille de sous-ensembles de X , contenant l'ensemble vide $\emptyset$, l'ensemble entier X , stable par intersections finies et stable par unions arbitraires. Les sous ensembles $U \in \tau$ sont dits *ouverts*, tandis que les ensembles F tels que $X \setminus F \in \tau$ sont dits *fermés*. Un exemple de topologie est l'espace des nombres réels $\mathbb{R}$ muni de sa topologie usuelle.

Au sein des espaces topologiques, les *espaces Noëthériens* sont les espaces (X, τ) dans lesquels il n'existe pas de séquence infinie strictement croissante d'ouverts. De manière équivalente, un espace (X, τ) est Noëthérien lorsque tout sous-ensemble $E \subseteq X$ est compact, notion topologique que je ne vais pas définir, mais qui est une manière topologique de généraliser les ensembles finis. Les espaces Noëthériens généralisent la notion de beau préordre, et peuvent être utilisés afin de démontrer la terminaison de certains programmes. Ces espaces ont été utilisés pour vérifier des systèmes de transitions dont la description topologique est plus adaptée qu'une présentation en terme d'ordres partiels, par exemple en utilisant la topologie de Zariski pour vérifier des programmes concurrents effectuant des calculs polynômiaux [Gou10].

Invariants Ordinaux. Dans le cas des beaux préordres comme des espaces Noëthériens, on peut mesurer la « complexité » intrinsèque de ces objets en utilisant des mesures *ordinales*. Pour simplifier, je vais seulement les définir dans le cas des beaux préordres, mais elles sont également définissables pour les espaces Noëthériens. Ainsi, DE JONGH et PARIKH [DP77] ont introduit le *type d'ordre maximal* comme le type d'ordre d'une linéarisation maximale, suivis par SCHMIDT [Sch81] qui introduit la notion de *hauteur ordinale* via la notion de chaîne croissante maximale, et par KŘÍŽ et THOMAS [KT90] qui ajoutent aux *invariants ordinaux* la *largeur ordinale*, uniformisent les définitions des trois invariants ordinaux et prouvent les premiers résultats arithmétiques qui les relient. Ces invariants ordinaux ont permis, dans le cas des WSTS, de fournir des bornes de complexité pour les algorithmes de vérification [Bon+13; Fig+11b; AFS15; Sch19a; Bal20].

Cependant, le calcul explicite ou l'obtention de bornes fines sur les invariants ordinaux d'un beau préordre s'avèrent souvent laborieuses et non triviales. En pratique, les beaux préordres utilisés en vérification sont souvent obtenus en combinant des beaux préordres plus simples, par des constructions algébriques (somme, produit cartésien, mots finis, ensemble des parties, etc). Comme les invariants ordinaux sont connus sur des beaux préordres « simples » (par exemple $\mathbb{N}$), l'effort de recherche a principalement été dirigé vers l'obtention formules permettant de calculer les invariants ordinaux d'un beau préordre composé (par exemple, calculer les invariants de $\mathbb{N} \times \mathbb{N}$ à partir de ceux de $\mathbb{N}$) [DP77; Sch20; VRW15; Wei09; AB99]. Cela permet de calculer en pratique les invariants ordinaux qui apparaissent dans les algorithmes de vérification. Récemment, DŽAMONJA, SCHMITZ et SCHNOEBELEN [DSS20a] ont écrit une revue de littérature sur le sujet en y ajoutant des résultats originaux au sujet des produits lexicographiques et des hauteurs ordinales des mots finis, multi-ensembles et arbres finis.

3.2.1 De nouveaux espaces

En collaboration avec Simon HALFON et Jean GOUBAULT-LARRECQ, j'ai démontré que certaines topologies sur les espaces de mots *transfinis* (c'est-à-dire, les fonctions de α vers X , où α est un ordinal fixé) sont Noëthériennes. Ces résultats sont intéressants car l'ensemble des mots infinis (donc, avec $\alpha = \omega$) muni de le préordre sous-mot éparc n'est déjà pas un beau préordre. Ainsi, de même que l'approche topologique permet d'éviter le problème de l'ensemble des parties, il permet aussi de considérer des mots ordinaux. En plus de démontrer que l'ensemble des mots *transfinis* pouvait être muni d'une topologie Noëthérienne, nous avons également calculé précisément les invariants ordinaux de ces espaces. Ces résultats ont été publiés à Colloquium Mathematicum [GHL23].

L'étude de ces espaces topologiques a été l'occasion pour moi de questionner la pertinence des topologies (mais aussi des ordres) placés sur les constructions classiques que l'on peut résumer ainsi : pourquoi le préordre sous mot éparc plutôt qu'un autre ? Cette question avait déjà été abordée par HASEGAWA [Has02], en utilisant des outils de théorie des catégories, mais possédait de sévères limitations. D'une part, la théorie semblait difficilement applicable à des cas concrets. D'autre part, elle ne permettait que de traiter des constructions *inductives* comme l'ensemble des mots finis, ou des arbres finis, mais pas des constructions comme l'ensemble des mots *transfinis*. Enfin, étant antérieure à l'utilisation des espaces Noëthériens en vérification, elle ne permettait pas de traiter le cas topologique. Dans mon article présenté à FoSSaCS 2023, j'ai développé une théorie qui répond à toutes ces limitations, et qui généralise strictement les travaux d'HASEGAWA. Au cœur de ce résultat, on trouve un théorème de *point fixe* pour les topologies Noëthériennes qui abstrait avantageusement les preuves *ad-hoc* de la littérature, basées sur des variations autour du schéma de preuve dit de la *mauvaise séquence minimale* de NASH-WILLIAMS [Nas65], souvent compliquées et parfois incorrectes [Gal97; Gou22b]. Ces théorèmes génériques ont permis d'introduire de nouveaux espaces Noëthériens et de retrouver les topologies *ad-hoc* introduites précédemment dans la littérature. Ces résultats ont été publiés à FoSSaCS 2023

[Lop23a].

3.2.2 L'ensemble des parties finies

On collaboration avec Simon HALFON, Philippe SCHNOEBELEN et Isa VIALARD, nous avons étudié les invariants ordinaux de l'ensemble des parties finies d'un beau préordre. Jusqu'alors, il n'existait essentiellement aucun résultat non-trivial sur le sujet. Dans un premier temps, nous avons montré formellement des bornes supérieures et inférieures « faciles », puis montré que celles-ci sont atteintes par certaines familles de beaux préordres. Les bornes supérieures et inférieures étant atteintes, et très éloignées les unes des autres, on déduit en particulier que ces invariants ordinaux ne peuvent pas être fonctionnels (se calculer uniquement à partir des invariants ordinaux des arguments du constructeur). Cependant, en restreignant les constructeurs autorisés (en interdisant de construire des beaux préordres avec une somme lexicographique), nous avons démontré que les invariants ordinaux des parties finies et du produit cartésien deviennent *fonctionnelles*, sous réserve d'adjoindre à la liste des invariants ordinaux leurs variantes « approximées ». Ainsi, nous avons défini une notion de « hauteur approximée », de « largeur approximée », et de « type d'ordre maximal approximé », qui en plus de permettre le calcul d'invariants par induction sur l'expression, semblent être intéressants en eux-mêmes [Abr+23].

3.3 Théorie des modèles finis

Cette partie est principalement basée sur deux publications [Lop21; Lop22]. En sus des présentations aux conférences où ces papiers ont été acceptés, j'ai eu l'occasion de présenter ces travaux au LIRMM (Université de Montpellier) et à l'Université de Varsovie en 2023, ainsi qu'au GT DAAL à Lille en 2022. Plus récemment, les résultats de ces travaux ont été récompensés par deux prix de thèse : le prix de thèse E. W. Beth 2024 ainsi que le prix Ackermann 2024.

Théorèmes de préservation. En théorie des modèles (classique), il existe une famille de théorèmes établissant la correspondance entre des préordres sur les structures relationnelles et des fragments de la logique du premier ordre. Ces théorèmes dits *théorèmes de préservation* sont construits autour de la notion de formule préservée par un préordre $\leq$, définie comme suit : une formule close φ est *préservée par* un préordre $\leq$ si l'ensemble des modèles de φ est clos par le haut pour $\leq$, c'est-à-dire, si $A \models \varphi$ et $A \leq B$, alors $B \models \varphi$. Un théorème de préservation pour un fragment $F \subseteq FO$ et un préordre $\leq$ a la forme suivante [CK90] :

Toute formule close $\varphi \in FO$ qui est *préservée par* le préordre $\leq$ est équivalente à une formule dans le fragment $F \subseteq FO$.

Par exemple, le théorème de préservation par homomorphisme indique que les formules φ qui sont préservées par homomorphismes de structures sont précisément celles que l'on peut écrire de manière *existentielle positive*. Cependant, ces résultats ne sont valides que dans le cas de la théorie des modèles classique, et leur relativisation à des classes de structures finies n'est pas triviale [ADGo8; Roso8]. En fait, il y a toute une littérature de résultats décrivant des exemples et contre-exemples de relativisations pour divers théorèmes de préservations et diverses classes de structures relationnelles finies [voir Tai59; AG87; AG94; Sto95; DS21; CF21; Kup21, pour des résultats négatifs] [voir Din92; ADKo6; ADGo8; Roso8; DRT10; SAC14; SAC16; BC19, pour des résultats positifs].²

3.3.1 Espaces préspectraux logiquement présentés

Comme expliqué dans l'introduction de ce rapport sur mes travaux de recherche, j'ai adapté les techniques basées sur les beaux préordres et les espaces Noëthériens à des classes de systèmes infinies dont l'espace des états est une classe de structures relationnelles finies. Une des idées a été d'adapter la notion d'espace Noëthérien à un cadre logique, en passant d'un espace (X, τ) à un triplet $(X, \tau, \mathcal{B})$ où $\mathcal{B}$ est une sous-algèbre Booléenne de $\mathcal{P}(X)$, c'est-à-dire, une famille de sous-ensembles de X stable par complémentaire, intersections finies et unions finies. Ces espaces, que j'ai appelé *espaces préspectraux logiquement présentés* (LPPS), généralisent les espaces Noëthériens à un cadre logique en prenant en compte l'interaction entre les sous-ensembles *ouverts* (topologie) et *définissables* (logique). Le nom de ces espaces vient de la notion d'espace spectral [elle même généralisant les espaces Noëthériens, voir DST19].

². Il semble exister une controverse sur l'usage de telles citations dans un document. Dans ce cas particulier, les énoncés précis des résultats sont moins importants que leur *nombre* et la *diversité* des auteurs impliqués.

L'exemple typique est celui d'une classe de structures relationnelles finies $\mathcal{C}$ munie de la topologie $\tau_{\subseteq_i}$, la topologie associée au préordre sous-structure induite $\subseteq_i$, et de la sous-algèbre Booléenne $\mathcal{B}_{FO}$ des ensembles définissables en logique du premier ordre dans $\mathcal{C}$. Dans ce cas, l'espace $(\mathcal{C}, \tau_{\subseteq_i}, \mathcal{B}_{FO})$ est un espace préspectral logiquement présenté si et seulement si :

1. Tout ouvert définissable $U \in \tau_{\subseteq_i} \cap \mathcal{B}_{FO}$ est *compact*, ce qui se traduit dans ce cas par le fait que tout ensemble ouvert définissable possède un nombre fini de modèles minimaux.
2. Tout ouvert est une union arbitraire d'ensemble définissables, ce qui est le cas par définition de la topologie $\tau_{\subseteq_i}$.

La condition (1) est une restriction de la notion d'espace Noëthérien (qui demande à ce que *tous les ouverts* soient compacts), tandis que la condition (2) est une condition de compatibilité entre la logique utilisée et la topologie d'intérêt. Le lien avec la notion de théorème de préservation se fait par la correspondance entre « avoir un nombre fini de modèles minimaux » et « être définissable dans une certaine logique ».

Dans le cas des beaux préordres et des espaces Noëthériens, le fait que ces propriétés soient difficiles à obtenir est compensé par l'existence d'une liste de constructeurs qui permettent d'obtenir des espaces complexes à partir d'espaces plus simples. Ainsi, les beaux préordres sont laissés stables par les opérations comme produit cartésien, union disjointe, mots finis, arbres finis, ensemble des parties finies et bien d'autres. J'ai démontré que les espaces préspectraux logiquement présentés capturent un sous ensemble (très large) des théorèmes de préservation qui relativisent au cas fini, et que ces espaces peuvent se composer de multiples manières (somme, produit cartésien, mots finis, produit en couronne), ce qui se prouve en utilisant à la fois des théorèmes topologiques (côté préspectral) et des théorèmes sur la logique au premier ordre (côté logiquement présenté). En un sens, ces espaces capturent les « bons » théorèmes de préservation et excluent ceux qui ne peuvent pas se composer de manière topologique.

La contribution est donc double : d'une part, j'ai introduit un objet intéressant qui permet de généraliser les beaux préordres et les espaces Noëthériens à un cadre logique, et d'autre part, j'ai montré que ces objets ont une application directe à la théorie des modèles finis en permettant de *composer* des résultats existants, ce qui était jusqu'alors une question ouverte et mal comprise.

3.3.2 Localité

Traditionnellement, la relativisation de théorèmes de préservation à des classes de structures finies repose sur une compréhension de la combinatoire de la logique du premier ordre sur des classes de structures ayant de fortes propriétés structurelles, par exemple avec un degré borné, ou des classes structures dites *éparses* [ADGo8; Roso8; Daw10]. Ces théorèmes utilisent très souvent comme première étape la localité de la logique du premier ordre, via le théorème de Gaifman [Gai82] permettant de réduire l'évaluation d'une formule sur une structure à l'évaluation de nouvelles formules sur des *voisinages* de celle-ci. Une fois cette étape effectuée, une analyse de la combinatoire des voisinages possibles permet ensuite de prouver qu'un certain théorème de préservation relativise. Cependant, ces deux étapes sont très souvent entremêlées et interdépendantes [ADGo8]. Par ailleurs, les résultats basés sur cette technique pour le *théorème de préservation par homomorphisme* [ADKo6; Daw10] se sont avérés faux [DE24], illustrant s'il le fallait la complexité de ces preuves.

Afin de mieux comprendre le comportement des espaces préspectraux logiquement présentés, j'ai étudié comment les théorèmes de *localité* de la logique du premier ordre pouvaient servir pour construire des théorèmes de préservation en prenant une forme de *limite*. Cette approche est, en quelque sorte, similaire à celle utilisée dans mes travaux sur les espaces Noëthériens [Lop23a]. Ma contribution principale a été de proposer un *découplage* entre l'analyse basée sur la logique (localité, combinatoire des voisinages) et l'analyse ultérieure basée *sur les voisinages* qui permet de conclure à un théorème de préservation.

La *localisation* d'une classe $\mathcal{C}$ est la famille dirigée d'ensembles $\mathcal{C}_{r,k}$ pour $r, k \in \mathbb{N}$, définis en considérant les r -voisinages autour de k points de structures dans $\mathcal{C}$. Par exemple, si $\mathcal{C}$ est la classe des *cycles* finis, alors $\mathcal{C}_{r,1}$ est la classe des cycles finis et des chemins finis de taille au plus $2r + 1$. Lorsqu'une classe $\mathcal{C}$ est héréditaire, c'est-à-dire close par sous-structure induite, les classes $\mathcal{C}_{r,k}$ sont incluses dans $\mathcal{C}$, et on a l'égalité $\mathcal{C} = \bigcup_{r,k} \mathcal{C}_{r,k}$. Ma principale motivation a été de montrer comment le théorème de localité de Gaifman permet de conclure à un théorème de préservation pour $\mathcal{C}$, simplement en le démontrant pour chaque $\mathcal{C}_{r,k}$, soit une forme de passage à la limite.

Pour cette étude, je me suis restreint à l'étude d'un unique théorème de préservation, le théorème de Łoś-Tarski, en me basant sur une étude préalable de Atserias, Dawar et Grohe [ADGo8], qui démontre par un argument combinatoire que le théorème de Łoś-Tarski relativise à toute classe de structures finies qui est héréditaire, close par union disjointe et de degré borné. Dans ce

cadre, j'ai démontré le théorème *local-vers-global* suivant, qui caractérise exactement les classes héréditaires et closes par union disjointe où le théorème de Łoś-Tarski relativise : pour de telles classes $\mathcal{C}$, le théorème de Łoś-Tarski relativise si et seulement si ce dernier relativise à toutes les *localisations* de $\mathcal{C}$.

Ce théorème permet d'obtenir de nouvelles classes, au-delà des classes de degré borné, sur lesquelles le théorème de Łoś-Tarski relativise. Par ailleurs, le résultat est obtenu par une méthode intéressante en elle-même, qui passe par une forme *positive* du théorème de localité de Gaifman. Cette version *positive* correspond à un fragment des formules au premier ordre, les formules *existentielles locales*, qui une fois paramétrées par le rayon r de localité autorisé, donnent une manière de quantifier le passage des formules *existentielles* (0-locales) aux formules *arbitraires* (∞ -locales).

Ces résultats ont été publiés à LICS en 2022 [Lop22].

3.4 Théorie des automates et transductions

Cette partie est principalement basée sur une publication à LICS 2023 [CDL23] et une pré-publication acceptée à STACS 2025 [Lop25]. Elle décrit des travaux que j'avais entrepris de manière indépendante durant ma thèse, mais qui occupent désormais une place centrale dans mon post-doctorat. Les résultats de [Lop25] ont été présentés au séminaire de l'équipe INRIA LINKS à Lille et au séminaire de l'équipe MOVE à Marseille en 2024. Ceux de [CDL23] ont été présentés au séminaire de l'équipe MOVE à Marseille en 2023.

Automates et transductions. Les automates finis sont des machines abstraites qui reconnaissent des langages, c'est-à-dire des ensembles de mots. La théorie des *transductions* vise à généraliser les résultats obtenus sur ces langages à des *fonctions* c'est-à-dire, passer de $L : \Sigma^* \rightarrow \mathbb{B}$ à $f : \Sigma^* \rightarrow \Gamma^*$, où $\mathbb{B}$ est l'ensemble des Booléens et Σ, Γ sont deux alphabets finis. La multitude de modèles de calculs équivalents pour les langages réguliers ne le sont plus lorsque ceux-ci se généralisent à des fonctions. Pour ces modèles de différentes expressivités, deux problèmes fondamentaux sont étudiés :

1. La décidabilité de l'équivalence : étant donné deux fonctions f et g , est-ce que $f = g$. Ce problème est connu comme décidable pour une grande partie des modèles de calculs, mais reste ouvert pour les fonctions polyrégulières, dont nous parlerons plus bas.
2. La décidabilité de l'apériodicité : étant donné une fonction f , est-ce que f est apériodique, c'est-à-dire, est-ce que f est équivalente à une fonction définissable en logique du premier ordre. Ce problème est aussi ouvert pour les fonctions polyrégulières.

La notion d'apériodicité est liée aux théorèmes fondamentaux établissant l'équivalence entre les langages *sans étoile*, les automates *sans compteur*, et la *logique du premier ordre* sur les mots finis [Sch65b; PP86; MP71; voir Str18, pour une vision historique]. Dans le cadre du mouvement général des *langages* vers les *fonctions*, la décidabilité de ce problème augure en général une bonne compréhension théorique et algorithmique du modèle de calcul considéré.

Automates pondérés et séries rationnelles. En parallèle du mouvement de fond allant des langages vers les fonctions [Sco67], une autre direction de recherche a été de généraliser les langages (qualitatifs) à des *langages pondérés* (quantitatifs). La notion d'automate pondéré (sur un semi-anneau) permet de considérer des fonctions $f : \Sigma^* \rightarrow \mathbb{S}$, c'est-à-dire des fonctions qui associent à chaque mot un *poids* au sein d'un semi-anneau $\mathbb{S}$. Les propriétés des automates pondérés sont très dépendantes de la structure de $\mathbb{S}$, et en général choisir $\mathbb{S}$ comme un *corps* permet d'appliquer les théorèmes d'algèbre linéaire et se placer dans un cadre bien compris. Deux questions ont retenu mon attention dans ce cadre :

1. Les théorèmes de « changement de semi-anneau », et plus particulièrement le problème de décision suivant : étant donné $f : \Sigma^* \rightarrow \mathbb{Q}$, est-ce que f peut-être réalisée par un automate pondéré utilisant des poids dans $\mathbb{N}$?
2. La définition d'une variante *apériodique* des automates pondérés. En effet, plusieurs définitions coexistent dans la littérature [DG19; Reu80], mais celles-ci ne généralisent pas correctement le concept d'apériodicité pour les langages réguliers.³

3. Pour les deux modèles cités, la fonction $w \mapsto (-1)^{|w|}$ est apériodique. Or, cela implique que la fonction indicatrice des mots de taille impaire est apériodique. Comme l'ensemble des mots de taille impaire est un langage *périodique*, cela montre que la notion d'apériodicité pour les fonctions et les langages ne coïncident pas.

Fonctions $\mathbb{N}$ -polyrégulières. Les fonctions polyrégulières forment un modèle de calcul permettant de modéliser un grand nombre de modifications *mots-à-mots* [Boj18]. Durant sa thèse de doctorat, Gaëtan DOUÉNEAU-TABOT a démontré que restreindre l’alphabet de sortie d’une fonction polyrégulière à un alphabet unaire permettait de simplifier de nombreux problèmes de décision [Dou21; Dou23]. Cette approche a aussi été prise par KIEFER, NGUYÊN et PRADIC [KNP23] pour étudier la vitesse de croissance des fonctions polyrégulières. La restriction à une sortie unaire permet de faire se rejoindre la théorie des fonctions polyrégulières, et la théorie des automates pondérés via la correspondance $(\mathbb{N}, +) \simeq (\{1\}^*, \cdot)$, et l’on parle alors de *fonctions $\mathbb{N}$ -polyrégulières* (NPoly) comme d’un sous ensemble des automates pondérés à coefficients dans $\mathbb{N}$. Ces fonctions avaient déjà été introduites indépendamment par SCHÜTZENBERGER [Sch62], et correspondent à une variante de la vision « quantitative » de MSO par KREUTZER et RIVEROS [KR13]. Le problème de l’équivalence est décidable pour les fonctions $\mathbb{N}$ -polyrégulières, puisque celui-ci est décidable pour la classe plus générale des automates pondérés à valeurs dans $\mathbb{Q}$ [BR11, Corollary 3.6].

3.4.1 Fonctions $\mathbb{Z}$ -polyrégulières

Même dans le cas des fonctions $\mathbb{N}$ -polyrégulières, le problème de l’apériodicité reste ouvert à ce jour. En collaboration avec Gaëtan DOUÉNEAU-TABOT et Thomas COLCOMBET, nous avons introduit une classe alternative de fonctions, les fonctions $\mathbb{Z}$ -polyrégulières ($\mathbb{Z}$ Poly), en autorisant des sorties négatives. Cette nouvelle classe de fonctions est contenue dans les fonctions calculables par des automates pondérés à valeurs dans $\mathbb{Z}$, mais ne forme plus une sous-classe des fonctions polyrégulières.

La première question que nous avons étudié est la possibilité de contrôler le degré de croissance de telles fonctions : pour une fonction f donnée, nous pouvons calculer le plus petit k tel que pour tout $w \in \Sigma^*$, $|f(w)| \leq C \times |w|^k$. Étant donné un tel k , nous pouvons aussi fournir une expression calculant f à l’aide de k boucles pour i de 1 à $|w|$ faire ... imbriquée, ce qui correspond à une forme d’optimisation de la fonction f . Ces résultats étaient déjà connus pour les fonctions à valeurs dans $\mathbb{N}$, mais leur adaptation à $\mathbb{Z}$ nécessitait une étude approfondie des potentielles « compensations » liées à l’autorisation de valeurs négatives. Au cœur de la preuve se trouve l’utilisation du théorème de factorisation de SIMON [Sim90a], qui nous a permis de décomposer une fonction polyrégulière sur les *mots* en une somme de fonctions très simples (de dénombrement) sur des *arbres*. Fort de ce premier résultat, nous avons pu construire une représentation *canonique* pour les fonctions $\mathbb{Z}$ -polyrégulières, sous la forme d’un automate des résiduels. À l’aide de cet objet canonique, nous avons élaboré une procédure de décision pour le problème de l’apériodicité, établissant par ailleurs de multiples caractérisations de celle-ci.

Ces travaux ont fait l’objet d’une publication à LICS 2023 [CDL23].

3.4.2 Fonctions $\mathbb{N}$ -polyrégulières

Notre objectif initial étant de décider l’apériodicité des fonctions polyrégulières, il faut constater que les résultats obtenus pour $\mathbb{Z}$ Poly ne sont pas directement applicables à NPoly. Cependant, nous avons conjecturé que $\text{NPoly} \cap \mathbb{Z}\text{SF} = \text{NSF}$ [Dou23, Conjecture 7.61]. Ainsi, il semblerait que comprendre le problème de l’appartenance de NPoly dans $\mathbb{Z}$ Poly, c’est-à-dire une version restreinte du problème équivalent pour les automates pondérés, serait une étape clé pour prouver $\text{NPoly} \cap \mathbb{Z}\text{SF} = \text{NSF}$, ce qui donnerait espoir pour la conception d’un algorithme pour le problème de décision de l’apériodicité dans le cas général.

Poursuivant cette intuition, j’ai étudié le problème de décision de NPoly au sein de $\mathbb{Z}$ Poly. Ce problème étant une restriction du problème équivalent pour les automates pondérés, j’ai d’abord lu les travaux de KARHUMÄKI [Kar77], qui a caractérisé quels *polynômes* sont calculables par des automates pondérés à valeurs dans $\mathbb{N}$. De manière plus générale, les fonctions qui m’intéressent sont les fonctions polyrégulières *commutatives*, dans le sens où $f(w)$ ne dépend que du multi-ensemble des lettres de w , et non de leur ordre. En généralisant ces résultats aux fonctions polyrégulières commutatives, j’ai découvert une erreur dans la preuve de KARHUMÄKI [Kar77]. Cela a été l’occasion de comprendre en détail le pouvoir expressif des fonctions $\mathbb{N}$ -polyrégulières sous l’hypothèse de commutativité. En particulier, j’ai démontré que sous cette hypothèse :

1. NPoly est une sous-classe décidable de $\mathbb{Z}$ Poly,
2. $\text{NPoly} \cap \mathbb{Z}\text{SF} = \text{NSF}$.

Ainsi, NSF est une sous classe décidable de NPoly, et les conjectures de DOUÉNEAU-TABOT [Dou23] sont confirmées dans le cas de fonctions commutatives. Ces résultats sont en cours de publication à STACS 2025 [Lop25].

3.5 Travaux en cours

Dans cette section, je liste mes travaux en cours, sans les détailler complètement.

Fonctions polyrégulières. Avec Antonio CASARES SANTOS et Mikołaj BOJAŃCZYK, nous étudions le problème de l'équivalence de fonctions polyrégulières, dans le cas particulier où leurs *entrées* sont unaires, c'est-à-dire, des fonctions de $\{1\}^* \rightarrow \Gamma^*$. Parallèlement, je travaille avec Omid YAGHOUBI, Antonio CASARES SANTOS et Mikołaj BOJAŃCZYK sur un modèle de calcul basé sur la complexité de la communication pour représenter des *fonctions* mot-vers-mots. C'est dans ce cadre que j'ai écrit un document sur les fonctions $\mathbb{N}$ -polyrégulières et les beaux préordres [Lop24a].

Vérification de programmes. En collaboration avec Rafał STEFAŃSKI, je développe un outil de vérification de programmes dans un sous-ensemble de Python utilisant la théorie des fonctions polyrégulières apériodiques. Ce travail est accessible sur GitHub [LS24].

Beaux préordres et langages. Suite à ma participation à l'événement Autoboz 2024 (dont je suis l'un des organisateurs), j'ai commencé à travailler avec Nathan LHOÏE et Lia SCHÜTZE sur la caractérisation des langages $L \subseteq \Sigma^*$ qui sont des beaux préordres pour le préordre *facteur*. En plus de fournir des caractérisations intéressantes, nous sommes capables de borner très finement les *invariants ordinaux* de ces langages. Les résultats préliminaires sont disponibles sur GitHub [LLS24].

Beaux préordres et transductions. Au début de ma thèse de doctorat, je me suis intéressé aux classes de structures relationnelles finies qui sont des beaux préordres pour l'ordre *sous-structure induite*, lorsque ceux-ci sont équipés d'un nombre fini (mais arbitraire) de couleurs. Cette notion dite de ω -beau préordre est au centre d'une conjecture de Pouzet : une classe de graphe est bien préordonnée pour le préordre sous graphe induit lorsqu'on lui adjoint *deux* couleurs l'est aussi pour n'importe quel nombre (fini) de couleurs [Pou72]. À cette époque, je n'avais pas les outils théoriques de théorie des automates pour approcher ce problème, et mes tentatives de généralisation des travaux de DALIGAULT, RAO et THOMASSÉ [DRT10] ont été infructueuses. Fort de mes nouvelles compétences (et en particulier du théorème de SIMON [Sim90a], ainsi que sa variante sur les arbres de COLCOMBET [Col07]), je m'emploie désormais à démontrer les conjectures suivantes :

1. Toute classe de graphes qui est un ω -beau préordre pour l'ordre *sous-structure induite* à largeur de clique bornée [DRT10, Conjecture 5].
2. Toute classe de graphes de largeur de clique bornée qui est un ω -beau préordre pour le préordre *sous-structure induite* est isomorphe à une classe d'arbres finis équipée d'une variante de l'ordre *gap embedding* de DERSHOWITZ et TZAMERET [DT03].

Dans le cas des graphes de *largeur de clique linéaire bornée*, j'ai déjà démontré cette seconde conjecture et établi la *décidabilité* d'être un ω -beau préordre [Lop24b]. Dans cette même prépublication, j'ai aussi démontré la seconde conjecture dans le cadre retenu par DALIGAULT, RAO et THOMASSÉ [DRT10]. Je travaille actuellement à généraliser ces résultats à la largeur de clique *non-linéaire*. Des discussions sur la première conjecture ont eu lieu avec Szymon TORUŃCZYK, même si aucun résultat n'a encore été obtenu. Les résultats partiels ont fait l'objet d'une présentation au séminaire de l'équipe AlGCO au LIRMM à Montpellier, et au séminaire de l'équipe LX au LaBRI à Bordeaux en 2024.

3.6 Rayonnement

En plus de mes publications dans des conférences internationales, j'ai été amené à participer au fonctionnement de la recherche et à sa diffusion en encadrant des étudiants, en faisant de la relecture d'articles pour des conférences de haut niveau (dont LICS, ICALP, CONCUR, MFCS, CSL, TOCL, LMCS). J'ai également été co-organisateur du workshop Autobóz 2024 [lien] et du séminaire Automates de l'IRIF de 2022 à 2023 [lien]. Enfin, j'ai effectué une journée d'animation d'informatique débranchée pour des élèves en primaire, dans le cadre de la fête de la science, à l'IRIF en octobre 2023, expérience que j'ai beaucoup appréciée, mais n'ai pas renouvelée depuis mon départ en Pologne.

Durant mon contrat post-doctoral, j'ai également été amené à encadrer deux étudiants, Arnav Garg et Ojas Maheshwari, sur un projet d'implémentation en rust de résultats théoriques en théorie des automates. Ce projet, hébergé sur un [dépôt GitHub public](#), a servi de base pour l'élaboration d'un logiciel de vérification de fonctions polyrégulières (cette fois développé en Haskell) qui est également disponible en ligne [sur GitHub](#).

Par ailleurs, j'ai participé à plusieurs événements et écoles sur invitation, listés ci-dessous :

- 2025-05-25 – 2025-05-30.** École d'hiver de recherche théorie des modèles finis. Les Houches, France. « *Finite and Algorithmic Model Theory* » [\[lien\]](#).
- 2025-03-30 – 2025-04-04.** École d'hiver de recherche sur les liens entre catégories et automates. Schloss Dagstuhl, Allemagne. « *Categories for Automata and Language Theory (25141)* » [\[lien\]](#).
- 2022-02-14 – 2022-02-18.** École d'hiver de recherche en logique, bases de données et complexité. CIRM, Luminy, France. « *Logique, bases de données et complexité* » [\[lien\]](#).
- 2023-01-16 – 2023-01-20.** École d'hiver de recherche en mathématiques discrètes et logique. CIRM, Luminy, France. « *Les mathématiques discrètes et la logique : des mathématiques à l'informatique* » [\[lien\]](#).
- 2022-03-28 – 2022-03-28.** Rencontre du groupe de travail DAAL, avec présentation des travaux de recherche. INRIA Lille, France. « *Rencontres GT DAAL* » [\[lien\]](#).
- 2023-02-21 – 2023-02-23.** Workshop sur les ordres bien quasi ordonnés. Université Claude Bernard, Lyon, France. « *WQO-BQO : What is up ?* » [\[lien\]](#).
- 2022-04-11 – 2022-04-12.** Réunion du projet ANR BraVAS avec présentation des travaux de recherche. LMF, Paris-Saclay, France. « *ANR BraVAS Project Meeting* ».
- 2022-01-31 – 2022-02-04.** École d'hiver de recherche théorie des modèles finis. Schloss Dagstuhl, Allemagne. « *Finite and algorithmic model theory (22051)* » [\[lien\]](#).

Méta-recherche. En dehors des activités stricto sensu scientifiques, j'ai également participé à des activités de méta-recherche, notamment en co-crédant l'organisation IDeA (Initiative Docteurs et Administrations) et en participant à l'initiative **TCS4F** (Theoretical Computer Science for Future) qui vise à limiter l'impact environnemental de la recherche en informatique théorique notamment en ne faisant plus de voyages nécessitant de prendre l'avion.

Dans le cadre du projet docteurs et administration, j'ai donné les interventions suivantes :

- 2025-03-28** 10 ans après votre entrée à l'ENS.
Département informatique, ENS Paris-Saclay.
- 2022-05-18** Rejoindre l'administration publique avec un doctorat.
Département informatique, ENS Paris-Saclay.
- 2022-11-30** Table ronde « Carrières des docteurs ».
CARaDOC, Centrale Supélec, Gif-sur-Yvette.

4 Projet de Recherche & Intégration

Résumé. Cette section présente mon projet de recherche en informatique, lequel repose sur deux piliers théoriques : d’une part la logique monadique du second ordre (MSO) et d’autre part la théorie des beaux préordres. En utilisant les techniques développées dans ces deux domaines, je souhaite explorer deux directions de recherche en vérification formelle de systèmes manipulant des données structurées (comme des bases de données) ou semi-structurées (comme des documents JSON, ou des fichiers CSV).

La première direction est consacrée à la vérification de systèmes de transition centrés sur les données, c’est-à-dire de systèmes de transitions dont les états sont des structures relationnelles finies (par exemple, une base de données d’un serveur web). La seconde direction est consacrée à la vérification de fonctions transformant des données semi-structurées (par exemple, des documents au format JSON), en poursuivant l’étude des fonctions polyrégulières. Ces deux directions se déclinent en différents axes de recherche, certains exploratoires et théoriques et d’autres plus appliqués. Chaque axe de recherche est lui-même découpé en jalons, qui sont des étapes concrètes et observables du projet permettant d’en mesurer l’avancement et de formuler précisément les théorèmes à démontrer ainsi que les logiciels à produire.

La cohérence de mon projet de recherche est double. D’une part, les deux directions de recherche sont complémentaires dans leur objet d’étude (systèmes évoluant *au long cours* et transformations *ponctuelles* de données). D’autre part, elles se rejoignent aussi dans les méthodes théoriques utilisées, principalement la logique monadique du second ordre, qui servira comme langage de *spécification* à la fois pour les systèmes de transitions (décrire l’espace des états, et les transitions possibles), et à la fois pour les fonctions transformant des données (via la notion de fonction polyrégulière). Un outil central lié à cette description est le théorème de factorisation de SIMON [Sim90b], et sa variante sur les arbres par COLCOMBET [Col07], qui fournissent des outils combinatoires essentiels à la réalisation des deux directions de ce projet.

4.1 Introduction

Les bases de données sont au cœur d’une grande majorité des systèmes informatiques modernes, que ce soit sous la forme de bases de données relationnelles, de base NoSQL, ou de fichiers structurés (CSV, JSON). Le traitement, le stockage, et la restitution de données structurées ou semi-structurées sont des enjeux majeurs pour les entreprises et les institutions publiques dans un contexte où la « donnée » est devenue une ressource stratégique,⁴ notamment pour l’apprentissage automatique. Ainsi, la certification et la vérification des systèmes manipulant des données prend une importance croissante. Un premier cas d’usage est l’obtention de garanties sur l’état d’un système *durant son utilisation*. Par exemple, lorsqu’une application en ligne expose une API pour accéder et modifier des données, ce qui est le cas de la majorité des sites internet. Dans ce cas d’usage, on souhaite pouvoir *certifier* que le système ne peut pas atteindre un état dit *d’erreur*, et ce quelle que soit la séquence d’opérations effectuées par les utilisateurs. Un second cas d’usage est la vérification des propriétés d’une transformation *ponctuelle*, typiquement lors d’une ingestion de données type « *ETL* » (Extract, Transform, Load), consistant à récupérer des données hétérogènes, les nettoyer, les convertir, et les insérer dans une base existante. Un exemple typique d’une telle transformation est la récupération de multiples jeux de données ouvertes (par exemple sur le site <https://www.data.gouv.fr>) et leur fusion dans un modèle unique. Cela pose des problèmes d’harmonisation des noms de champs, de formats d’entrée, et plus généralement de cohérence des données. Pour fixer un exemple qui servira de modèle dans ce projet, c’est précisément ce que fait la bibliothèque logicielle *french-cities* développée par la DREAL Hauts-de-France, qui permet de nettoyer des jeux de données mentionnant des villes françaises [Gra24].

4.1.1 Direction 1 : Systèmes centrés sur les données

Dans ce projet, les données structurées seront représentées mathématiquement par des *structures relationnelles finies*. Dans les exemples, nous utiliserons les *graphes finis coloriés*, qui sont des structures finies munies d’une relation binaire d’adjacence et de multiples prédicats unaires de couleurs. Pour modéliser les requêtes sur les structures finies, nous utiliserons des formules de la *logique du premier ordre* (FO), qui correspond précisément au calcul relationnel, lui-même formant un fragment raisonnable⁵

4. La quantité et l’hétérogénéité des données à traiter rend la tâche difficile, d’où une demande croissante pour les *data scientists* que l’on peut observer sur les offres d’emploi de la fonction publique <https://www.data.gouv.fr/fr/datasets/les-offres-diffusees-sur-choisir-le-service-public>

5. Le langage SQL étant *Turing-complet*, il n’est pas adapté à la vérification formelle.

du langage de requêtes SQL [Cod70]. On pourra par exemple demander s'il existe dans un graphe G un sommet ayant degré 3 et dont tous les voisins sont de couleur rouge.

Lors de la vérification d'un système informatique, un des problèmes fondamentaux est de s'assurer que le système ne se trouve jamais dans un état *invalide*. Ceci est mathématiquement formalisé en considérant l'ensemble $\mathcal{C}$ des états possibles du système, une relation de transition $\rightarrow \subseteq \mathcal{C} \times \mathcal{C}$ décrivant les évolutions possibles du système, ainsi qu'un ensemble $\text{Err} \subseteq \mathcal{C}$ d'états invalides. La question de la vérification est alors de décider si, pour tout état initial $c_0 \in \mathcal{C}$, le système ne peut jamais atteindre un état invalide (dans Err) en suivant les transitions de $\rightarrow$. Lorsque le système est *fini*, c'est-à-dire que l'ensemble $\mathcal{C}$ est fini, la décision de l'*accessibilité* d'un état cible t depuis un état initial i est relativement simple, mais ce n'est pas le cas qui nous intéresse dans ce projet. En effet, nous allons nous intéresser au cas où $\mathcal{C}$ est une classe (infinie) de structures finies, représentant les états possibles de notre base de données en production. En revanche, la relation de transition $\rightarrow$ d'un système ne sera pas arbitraire : elle sera supposée *compatible avec la logique*.⁶ Plus précisément, lorsque X est un ensemble FO-définissable de $\mathcal{C}$, on supposera que l'ensemble $\rightarrow^{-1}(X)$ — c'est-à-dire, l'ensemble des éléments de $\mathcal{C}$ dont une étape de transition amène à X — est lui-même FO-définissable.⁷ Lorsque l'ensemble des états est une classe $\mathcal{C}$ de structures finies, et que la relation de transition $\rightarrow$ est compatible avec la logique, on dira que $(\mathcal{C}, \rightarrow)$ est un *système centré sur les données*.

Pour illustrer notre problématique, prenons comme *exemple de système* la classe des *graphes* finis excluant le sous-graphe induit P_4 (un chemin de longueur 4). Sur cette classe, on définit la relation $G \rightarrow_{\text{ex}} G'$ lorsque G' est obtenu en sélectionnant 5 sommets indépendants dans G et en ajoutant un nouveau sommet relié à ces 5 sommets. Il semble a priori difficile de décider si, étant donné un ensemble Err_{ex} d'états invalides et graphe initial $G_i \in \mathcal{G}_{\neg 4}$, on peut atteindre un état d'erreur en suivant les transitions de $\rightarrow_{\text{ex}}$. Cependant, cette question est en fait *décidable* dès que Err_{ex} est défini via un ensemble fini de *sous-graphes interdits*, par exemple, Err_{ex} est l'ensemble des graphes contenant un triangle.

En effet, on peut démontrer que la classe $\mathcal{G}_{\neg 4}$, équipée du préordre $G \subseteq_i H$ si G est un *sous-graphe induit* de H ,⁸ est ce qu'on appelle un *beau préordre* (WQO). Cela signifie que toute séquence $(G_i)_{i \in \mathbb{N}}$ de graphes de $\mathcal{G}_{\neg 4}$ possède une *paire croissante*, $G_i \subseteq_i G_j$ pour $i < j$. Par ailleurs, il est possible de prouver que la relation $\rightarrow_{\text{ex}}$ est *compatible* avec l'ordre $\subseteq_i$, car si $G \rightarrow_{\text{ex}} H$ et $G \subseteq_i G'$, alors il existe $H' \in \mathcal{G}_{\neg 4}$ tel que $H \rightarrow_{\text{ex}} H'$ et $G' \subseteq_i H'$. De plus, l'ensemble Err_{ex} est un ensemble *clos par le haut* pour l'ordre $\subseteq_i$ sur $\mathcal{G}_{\neg 4}$, c'est-à-dire, si pour tout graphes $G, H \in \mathcal{G}_{\neg 4}$, $G \in \text{Err}_{\text{ex}}$ et $G \subseteq_i H$ impliquent $H \in \text{Err}_{\text{ex}}$. Ainsi, le triplet $(\mathcal{G}_{\neg 4}, \subseteq_i, \rightarrow_{\text{ex}})$ est un *système de transition bien structuré* (WSTS), qui sont des systèmes pour lesquels l'accessibilité d'un ensemble clos par le haut depuis un état initial est *décidable*.

Le formalisme des systèmes de transition bien structurés a précisément été introduit pour la vérification de systèmes de transitions sur un ensemble infini d'états possibles [Abd+96]. Il a été utilisé avec succès pour résoudre de nombreux problèmes [FS01], et son étude théorique a été récompensée par le prix CAV 2017 pour des contributions fondamentales à la vérification formelle assistée par ordinateur [Abd+17]. Malheureusement, toutes les classes $\mathcal{C}$ de structures finies ne sont pas des beaux préordres (et d'ailleurs, pour quels préordres ?), ce qui pose un frein majeur à l'utilisation générique de ces techniques. De plus, même dans le cas où $(\mathcal{C}, \leq)$ est un beau préordre, il reste à vérifier que la relation de transition $\rightarrow$ est compatible avec l'ordre $\leq$. Enfin, pour appliquer des algorithmes de vérification il faudra construire des représentations satisfaisant les hypothèses d'effectivité des algorithmes de vérification de type WSTS. C'est ce qui motive ma première direction de recherche.

Direction 1. Développer des techniques de vérification pour les systèmes centrés sur les données en utilisant la théorie des WSTS.

4.1.2 Direction 2 : Vérification de transformations

Pour étudier des *transformations de données*, comme celles décrites par les pipelines ETL, la notion mathématique retenue dans ce projet est celle de *transductions* entre structures finies, c'est-à-dire, des fonctions (ou relations) définies à l'aide de formules logiques. Dans le cas restreint où les données d'entrée sont des chaînes de caractères, cela correspond à l'étude des *transductions mot-vers-mot* qui est un domaine très actif en théorie des automates [voir la récente synthèse de MP19]. Dans ce cadre, la logique

6. En effet, les modifications de la base de données se font *en pratique* dans le même langage que celui qui est utilisé pour les requêtes, c'est-à-dire SQL, et dans notre cas, FO.

7. La définition s'inspire de ce qui est fait pour des systèmes de transitions $(X, \rightarrow)$ où X est un espace *topologique* [Gou07; Gou10].

8. C'est-à-dire, il existe un sous-ensemble $V' \subseteq V(H)$ tel que $G = (V', E(H) \cap (V' \times V'))$.

du premier ordre (FO) est souvent remplacée par la *logique monadique du second ordre* (MSO) qui est plus expressive et correspond mieux aux *modèles de calculs* utilisés en pratique. Par exemple, les automates finis déterministes calculent exactement les propriétés définissables en MSO sur les mots finis, et il faut restreindre ce modèle de calcul pour obtenir uniquement les propriétés définissables au premier ordre.

Un de ces modèles de calcul, celui des *fonctions polyrégulières* [Boj18] fournit un modèle de transductions *mot-vers-mot* suffisamment riche pour y encoder des transformations de données structurées tout en maintenant de bonnes propriétés théoriques. En pratique, une grande partie des *pipelines* de type ETL font des modifications sur des données semi-structurées en utilisant un fragment très restreint du langage hôte (par exemple, un sous ensemble du langage Python dans le cas de *french-cities*). Prenons comme exemple une implémentation volontairement simplifiée d'une fonction `enrich_row_city` qui prend en argument un dictionnaire (une ligne de donnée) et renvoie un dictionnaire *enrichi* avec des informations spécifiques à la commune dont la ligne contient le code postal, que l'on décrit dans l'Algorithme 1.

Le code de l'Algorithme 1 ne contient pas de *boucles tant que* (While), ne manipule pas d'entiers, et les dictionnaires sont de profondeur finie égale à 1. En d'autres termes, la fonction `enrich_row_city`, lorsque la variable globale `sanitized` est fixée, est une fonction polyrégulière sous réserve d'encoder un dictionnaire comme une liste de couples clef-valeur, sérialisée en chaîne de caractères (par exemple, au format JSON). D'ailleurs, dans la suite du projet, on utilisera implicitement cette conversion pour parler de fonctions manipulant des dictionnaires comme des fonctions polyrégulières sur les mots.

Algorithme 1 : Code prototype d'un pipeline ETL pour les communes françaises.

```

1  def enrich_row_city(row : dict) -> dict:
2      """ Enrichit une ligne avec les données de la base de données des communes
3          stockées dans une variable globale `sanitized`.
4      """
5      global sanitized
6      # De multiples champs peuvent contenir le code postal
7      post_codes_fields = ["code_post", "code_postal", "code postal"]
8      for post_code_field in post_codes_fields:
9          # Le code postal n'est pas forcément indiqué dans chacun des champs
10         code_post = row.get(post_code_field, None)
11         # Le code postal peut contenir des caractères non numériques / des espaces
12         code_post_num = normalize_code_post(code_post) # fonction de normalisation
13         # On cherche la commune correspondante dans la base de données
14         sanitized_row = match_row(sanitized, code_post_num) # fonction de recherche
15         # Si une commune correspond
16         if sanitized_row is not None:
17             # On renvoie l'union des deux dictionnaires
18             return (row | sanitized_row)
19         # Si aucun code postal n'est trouvé, on renvoie la ligne inchangée
20     return row

```

Il est possible de *certifier* des transformations polyrégulières sous la forme de *triplets de Hoare* $\{\varphi\}P\{\psi\}$ où φ et ψ sont deux propriétés logiques, et P est une fonction polyrégulière. Le sens d'un tel triplet est que pour toute entrée du programme P satisfaisant φ , la sortie satisfait ψ . Par exemple, on peut certifier que la fonction `enrich_row_city` satisfait le triplet de Hoare $\{\top\}\text{enrich_row_city}\{\psi\}$ où $\top$ est la propriété vraie, et ψ est la propriété « le champ `département` s'il existe, est dans la liste des départements Français ». Ceci est possible car une fonction polyrégulière f est continue au sens où toute propriété ψ MSO-définissable sur les sorties de f peut se transformer en une propriété ψ_f sur les entrées de f satisfaisant : pour tout mot d'entrée w , $f(w) \models \psi \iff w \models \psi_f$ [Boj18, Théorème 1.7]. Ainsi, pour certifier le triplet précédent, il suffit de calculer ψ_f à partir de ψ et vérifier si ψ_f est vraie pour tous les mots d'entrée, ce qui est possible car ψ_f est calculable par un automate fini déterministe.

Il n'existe à ce jour aucun prototype permettant d'évaluer la viabilité de ces techniques sur des exemples pratiques. Par ailleurs, au-delà de la vérification de triplets de Hoare, une question théorique majeure reste ouverte : la décidabilité de l'équivalence de

deux fonctions polyrégulières. En effet, un grand nombre de certifications s'expriment sous la forme d'une équivalence entre deux fonctions; par exemple, on pourrait vérifier que sur des entrées ne contenant pas de code postal la fonction `enrich_row_city` de l'Algorithme 1 se comporte bien comme la fonction identité. C'est ce qui motive ma seconde direction de recherche.

Direction 2. *Développer des outils de vérification pour les fonctions polyrégulières.*

4.2 Direction 1 : systèmes centrés sur les données

La première condition pour appliquer des algorithmes de vérification de type WSTS est que la classe $\mathcal{C}$ de structures finies soit un beau préordre pour l'ordre choisi. Dans ce projet, on se limitera à *deux* préordres naturels sur les classes de structures finies $\mathcal{C}$: le préordre *sous-structure induite* ($\subseteq_i$) et le préordre *d'homomorphisme* ($\leq_{\text{hom}}$). Ces deux préordres sont intéressants pour des raisons différentes. D'une part, la notion de sous-structure induite est fondamentale en théorie des modèles finis, et d'autre part, la notion d'homomorphisme correspond précisément à une classe de requêtes FO-définissables sur les structures finies d'importance pratique en théorie des bases de données : les requêtes *UCQ* (Union of Conjunctive Queries), correspondant aux traditionnelles requêtes de la forme **SELECT** ... **FROM** ... **JOIN** ... **WHERE** ... [AHV95]. Ainsi, un objectif majeur de ce projet est de caractériser les classes $\mathcal{C}$ de structures finies qui sont des beaux préordres pour l'ordre $\subseteq_i$ ou $\leq_{\text{hom}}$. Cette étude s'inscrit dans une continuation de l'étude des systèmes décrits en logique de Presburger [FG19]. J'ai par ailleurs obtenu des résultats partiels sur des préordres encodés dans des ensembles de mots finis [LLS24], ou sur des classes de graphes ayant largeur de clique linéaire bornée [Lop24b].

Axe 1

Classier les classes de structures finies qui sont des beaux préordres pour l'ordre $\subseteq_i$ ou $\leq_{\text{hom}}$.

Historiquement, l'approche pour la vérification de systèmes infinis utilisant des systèmes de transition bien structurés a été de construire des espaces d'états $\mathcal{C}$ munis d'un beau préordre en utilisant des *constructeurs* : produits, unions, mots finis, arbres, etc [Dem+12; Hig52; Kru72; DT03]. Par exemple, un réseau de Petri avec un certain nombre k de places se représente en utilisant des k -uplets d'entiers, c'est-à-dire $\mathbb{N}^k$, qui est un beau préordre lorsque muni de l'ordre produit. Cependant, ces approches *bottom-up* se focalisant sur les constructeurs préservant les beaux préordres ne sont pas adaptées à la vérification de systèmes centrés sur les données, où la classe de structures $\mathcal{C}$ et l'ordre sur ces structures $\leq$ (avec $(\leq) \in \{\leq_{\text{hom}}, \subseteq_i\}$) forment une *entrée du problème*. En première approche, on se limitera à des classes de *graphes finis* en lieu et place des structures plus générales.

Dans le cas des graphes, deux propriétés plus fortes que celle de beau préordre sont déjà étudiées dans la littérature. Étant donné une classe de graphe $\mathcal{C}$, on peut se demander si elle est un beau préordre pour l'ordre $\subseteq_i$ *après* l'ajout de k couleurs finies sur les sommets, on dira alors que la classe est un *k-beau préordre*. Si la classe $\mathcal{C}$ est un k -beau préordre pour tout k , on dira que c'est un *ω -beau préordre*. Enfin, si la classe $\mathcal{C}$ est un beau préordre lorsque coloriée par n'importe quel autre beau préordre $(X, \leq)$, on dira que $\mathcal{C}$ est un *constructeur de beaux préordres*. Il a été conjecturé que toute classe de graphe $(\mathcal{G}, \subseteq_i)$ est un 2-beau préordre si et seulement si elle est un ω -beau préordre par Pouzet [Pou72], et je conjecture que l'équivalence s'étend à être un constructeur de beaux préordres. Caractériser les constructeurs de beaux préordres est un enjeu en soi, car cela permettrait d'expliquer la différence entre les ordres linéaires [Hig52] et les arbres [Kru72] (qui sont des constructeurs de beaux préordres) et d'autres classes comme celle de tous les graphes finis (qui n'est pas un constructeur de beaux préordres).

Afin de construire des *algorithmes* qui décident si une classe $\mathcal{C}$ est un beau préordre pour l'ordre $\subseteq_i$ ou $\leq_{\text{hom}}$, il faut convenir d'une représentation. Les classes seront spécifiées en logique monadique du second ordre, via une MSO-transduction depuis une classe d'arbres finis. Cette représentation décrit précisément⁹ des classes de structures de *largeur de clique bornée* et de *largeur de clique linéaire bornée* [CE12], inventées pour généraliser les langages réguliers aux classes de graphes. Par exemple, la classe $\mathcal{G}_{\neg 4}$ correspond à la classe des *cographes*, qui est la plus petite classe de graphes contenant le graphe à un seul sommet et close par les opérations de complémentation des arêtes et union disjointe. Ainsi, cette classe est immédiatement l'image via une MSO-transduction d'une classe d'arbres dont les nœuds internes sont étiquetés par les opérations d'union et de complémentation. Notre algorithme permettrait alors de répondre *automatiquement* à la question $(\mathcal{G}_{\neg 4}, \subseteq_i)$ est-il un beau préordre ?

⁹. Plus précisément, une classe de graphe $\mathcal{C}$ est de largeur de clique bornée si et seulement si elle est une *sous-classe* d'une classe de graphes définie par une MSO-transduction.

Jalon 1. *Démontrer que l'on peut décider si une classe $\mathcal{G}$ de graphes finis munie du préordre $\subseteq_i$ est un constructeur de beaux préordres, où $\mathcal{G}$ est donnée comme une MSO-transduction depuis la classe des arbres finis. Vérifier que cette classe est un ω -beau préordre si et seulement si elle est un constructeur de beaux préordres.*

J'ai déjà obtenu des résultats positifs pour les classes de largeur de clique linéaire bornée [Lop24b]. Au cœur des preuves de ces résultats se trouve le théorème de factorisation de SIMON [Sim90b], qui permet de faire le lien entre le préordre sous-structure induite sur la classe de graphe et le préordre de *gap embedding* sur les mots finis. Je m'emploie à compléter ces résultats pour la largeur de clique non-linéaire, à l'aide de la variante du théorème de SIMON sur les arbres de COLCOMBET [Col07].

Par ailleurs, il a été conjecturé que toute classe de graphe $(\mathcal{G}, \subseteq_i)$ qui est un 2-beau préordre a une largeur de clique bornée [DRT10, Conjecture 5]. Cette conjecture, au-delà de son intérêt théorique, validerait l'utilisation de MSO comme langage décrivant les classes de structures finies. C'est ce qui motive le jalon suivant.

Jalon 2. *Démontrer que toutes les classes de graphes finis $\mathcal{G}$ munies du préordre $\subseteq_i$ qui sont des ω -beaux préordres ont largeur de clique bornée.*

Une fois le problème bien compris sur les classes de graphes et l'ordre *sous-structure induite*, un enjeu majeur sera la généralisation de ces résultats dans trois directions : passer des graphes aux structures relationnelles, étudier le cas de l'ordre d'homomorphisme, et enfin, de ω -beau préordre à beau préordre. La faisabilité de cette dernière généralisation est justifiée par un premier résultat que j'ai obtenu, permettant de décider si un préordre (représenté comme un langage de mots) est un beau préordre pour l'ordre *facteur* [LLS24].

Jalon 3. *Démontrer que le problème suivant est décidable : étant donné une classe $(\mathcal{C}, \leq)$ de structures finies où $(\leq) \in \{\subseteq_i, \leq_{\text{hom}}\}$, et où $\mathcal{C}$ est donnée comme une MSO-transduction depuis la classe des arbres finis, est-ce que $(\mathcal{C}, \leq)$ est un beau préordre ?*

Axe 2

Développer une théorie permettant la vérification des systèmes centrés sur les données basée sur des systèmes de transition bien structurés.

Lorsque l'on considère un algorithme basé sur un beau préordre, la complexité de l'algorithme est profondément liée à la « complexité » du préordre sous-jacent [Sch17]. Une manière de mesurer cette complexité est de considérer ses *invariants ordinaux* : par exemple, le type d'ordre maximal (ou m.o.t.), défini à l'origine par DE JONGH et PARIKH [DP77], est le type d'ordre de la linéarisation maximale d'un ensemble bien préordonné. Dans le cas d'un ensemble fini, le m.o.t. est précisément la taille de l'ensemble. De meilleures bornes de complexité ont été obtenues en considérant deux autres paramètres [Sch19b] : la *hauteur ordinale* introduite par SCHMIDT [Sch81], et la *largeur ordinale* de KRÍŽ et THOMAS [KT90]. Par conséquent, lors de la caractérisation des classes de structures [bien préordonnées] pour l'ordre $(\subseteq_i)$ ou $(\leq_{\text{hom}})$, nous nous intéresserons également à calculer des bornes sur leurs invariants ordinaux.

Dans un cadre théorique différent, en représentant des ordres partiels $(X, \leq)$ par des langages $L \subseteq \Sigma^*$ ordonnés par l'ordre *facteur*, j'ai obtenu avec Nathan LHOTE et Lia SCHÜTZE des bornes extrêmement fines montrant que les langages *algébriques* de mots qui sont des beaux préordres pour l'ordre *infixe* ont une largeur ordinale strictement inférieure à ω^2 , et par conséquent un type d'ordre maximal strictement inférieur à ω^3 [LLS24]. En comparaison, l'ordre *sous-mot épars* sur Σ^* qui le rend bien préordonné a un type d'ordre maximal égal à $\omega^{|\Sigma|-1}$ [DSS20b, Corollaire 3.9, Théorème 4.21]. Ces résultats montrent que la complexité des beaux préordres *représentables* est significativement plus faible que celle des beaux préordres arbitraires.¹⁰

¹⁰. En effet, l'ensemble Σ^* avec l'ordre *sous-mot épars* est isomorphe (en tant qu'ordre) à un langage L de mots finis sur un alphabet binaire muni de l'ordre *facteur* [LLS24]. Ainsi, il existe des langages complexes, mais, il ne sont pas *représentables*.

Jalon 4. *Obtenir des bornes fines sur les invariants ordinaux des classes de structures finies qui sont des beaux préordres pour l'ordre $\subseteq_i$ ou $\leq_{\text{hom}}$.*

Afin de décider si un système $(\mathcal{C}, \rightarrow, \leq)$ est un WSTS, nous avons fixé la représentation de $\mathcal{C}$ et choisi $\leq$ comme étant $(\subseteq_i)$ ou $(\leq_{\text{hom}})$. Il reste à fixer une représentation de la relation de transition $\rightarrow$. À terme, il serait pertinent de considérer des transitions calculées par fonctions polyrégulières, mais ce modèle de calcul pose déjà des problèmes théoriques importants dans le cas d'une seule fonction, et il est difficile d'imaginer l'étude de compositions arbitraires de telles fonctions. Ainsi, dans un premier temps on peut se restreindre à la description la plus simple des transformations d'une structure A dans une structure B . Ainsi, on utilisera une formule $\varphi \in \text{MSO}$ sur les structures de la forme $A \uplus B$ ayant un prédicat binaire $H(x, y)$ pour représenter la relation $\rightarrow : A \rightarrow B$ si et seulement s'il existe une fonction h du domaine de A vers le domaine de B , telle que $(A \uplus B, \{(x, h(x)) \mid x \in A\})$ satisfait la formule φ . On dira d'une telle relation $\rightarrow$ qu'elle est **MSO-reconnaissable**. On peut ainsi décrire des relations comme « B est obtenue en ajoutant un sommet à A », ou « B est obtenue en fusionnant deux sommets de A ». En particulier, on peut reconnaître les relations $(\leq_{\text{hom}})$ et $(\subseteq_i)$ par une telle formule MSO.

Jalon 5. *Étant donné une classe $\mathcal{C}$ définie par une transduction MSO depuis la classe des arbres finis, un préordre $\leq \in \{\subseteq_i, \leq_{\text{hom}}\}$, et une relation $\rightarrow$ MSO-reconnaissable, décider si $(\mathcal{C}, \leq, \rightarrow)$ est un système de transition bien structuré.*

En pratique, les algorithmes de vérification de systèmes de transition bien structurés requièrent un certain nombre d'hypothèses d'effectivité (représentation des entrées, calculabilité des transitions, etc.). Ces questions d'effectivité ont fait l'objet d'une thèse de Doctorat de HALFON [Hal18], où il est démontré qu'il suffit pour bénéficier des résultats de décidabilité de pouvoir représenter les *idéaux* d'un beau préordre $(X, \leq)$, qui sont les sous ensembles $I \subseteq (X, \leq)$ simultanément *clos par le bas*¹¹ et *dirigés*¹² pour l'ordre $\leq$. Par exemple, dans l'ensemble $\mathbb{N}^2$ muni de l'ordre produit, les idéaux peuvent se décrire à l'aide de couples $(a, b) \in (\mathbb{N} \uplus \{\infty\})^2$, où (a, b) sert à représenter l'ensemble des couples (x, y) telles que $x \leq a$ et $y \leq b$, en prenant pour convention que ∞ est plus grand que tout élément de $\mathbb{N}$. Pour une classe de structures finies $\mathcal{C}$, munie de l'ordre $\subseteq_i$, obtenir une représentation *finie* des idéaux de $\mathcal{C}$ n'est pas évident, et il s'ajoute à cela des questions d'effectivité de cette éventuelle représentation : il faut pouvoir calculer l'intersection de deux idéaux, vérifier si un élément est dans un idéal, vérifier si un idéal est vide, etc.

Jalon 6. *Construire une représentation effective des idéaux pour des classes de structures $(\mathcal{C}, \subseteq_i)$, lorsque $\mathcal{C}$ est décrite par une MSO-transduction depuis une classe d'arbres finis.*

Axe 3

Établir une correspondance entre les algorithmes de requêtes sur des bases de données incomplètes, et les systèmes de transition bien structurés.

En théorie des bases de données, un problème classique est celui de l'évaluation de requêtes sur des bases de données incomplètes. Dans ce cadre, on dispose d'une base de données D et d'un ensemble de contraintes Σ sur la complétion de D . On cherche alors à savoir si toutes les complétions (potentiellement infinies) de D *valides* satisfont la requête Q . On notera $(D, \Sigma) \models_{\mathcal{C}} Q$ lorsque toutes les complétions de D au sein de la classe $\mathcal{C}$ qui satisfont Σ satisfont aussi Q . Un semi-algorithme est utilisé classiquement pour répondre à ce type de problèmes : le semi-algorithme de *Chase*. Notons que ce processus est plutôt une famille de semi-algorithmes dont il existe de nombreuses variantes, gardant cependant une structure commune [DNR08a].

Prendre en compte l'existence de *complétions infinies* ne correspond pas au cadre de travail de ce projet qui se concentre uniquement sur les structures finies. Dans le cas où les contraintes Σ ont la propriété de *finite controllability* [CGL12, Théorème 9], cette distinction n'a pas lieu d'être car, pour toute base de données D initiale, pour toute requête Q (UCQ), l'ensemble des réponses à une requête Q sur les complétions *finies* coïncide avec l'ensemble des réponses sur les complétions *infinies*. Pour simplifier, dans la suite de cette description, on se restreindra à des contraintes Σ qui ont cette propriété de *finite controllability*.

11. Un ensemble I est *clos par le bas* si pour tout $x \in I$ et tout $y \in X$ tel que $y \leq x$, on a $y \in I$.

12. Un ensemble I est *dirigé* s'il est non vide et que pour tout $x, y \in I$, il existe $z \in I$ tel que $z \leq x$ et $z \leq y$.

Prenons comme exemple une base de données qui est un graphe $G \in \mathcal{G}_{-4}$, une requête Q de type UCQ, et comme contrainte Σ la propriété qui vérifie si un graphe H satisfait « pour tout quintuplet de sommets distincts indépendants de H , il existe un sommet de H relié à ces cinq sommets »¹³ et « $H \in \mathcal{G}_{-4}$ ». La *Chase* (dans sa variante la plus simple), consiste à partir du graphe G , et vérifier si $G \models \Sigma$. Si oui, on peut répondre directement en évaluant Q sur G . Sinon, on peut trouver un quintuplet de sommets indépendants x_1, x_2, x_3, x_4, x_5 qui n'a pas de sommet relié à ces cinq sommets, et créer un nouveau graphe H obtenu en ajoutant un sommet relié à ces cinq sommets, sous réserve que cela ne crée pas un chemin de taille 4. On peut alors répéter l'opération sur H , en espérant que le processus termine.

Dans cet exemple, le système de transition $(\mathcal{G}_{-4}, \subseteq_i, \rightarrow_{\text{ex}})$ est un système de transition bien structuré, et $\rightarrow_{\text{ex}}$ correspond précisément aux transitions effectuées par le semi-algorithme de *Chase*. En fait, le semi-algorithme de *Chase* vérifie si, en partant de G , on peut atteindre un état d'erreur, c'est-à-dire une base de données qui satisfait $\Sigma \wedge \neg Q$. Si on note $\text{Reach}(G)$ l'ensemble des bases de données atteignables depuis G via la relation $\rightarrow_{\text{ex}}$, alors $(G, \Sigma) \models_{\mathcal{G}_{-4}} Q$ si et seulement si $\text{Reach}(G) \cap \downarrow(\Sigma \wedge \neg Q) = \emptyset$.

Cet algorithme est en apparence très semblable à un algorithme de vérification pour les systèmes de transition bien structurés. En effet, il est possible pour certains WSTS de calculer une approximation¹⁴ de l'ensemble $\text{Reach}(G)$ des états accessibles depuis G en utilisant un algorithme de saturation très similaire à celui de la *Chase* [FG20], qui trouve son origine dans les travaux de KARP et MILLER [KM69] pour les réseaux de Petri. Puisque ces deux semi-algorithmes sont très similaires, il est naturel de se demander si des résultats sur les classes de structures $\mathcal{C}$ obtenues dans les deux axes précédents peuvent être utilisés pour expliquer leur terminaison. Cet objectif de recherche est exploratoire : c'est le seul pour lequel aucun résultat préliminaire n'est disponible, mais il est motivé par une forte intuition que l'interaction entre logique et théorie de l'ordre peut apporter des résultats significatifs, comme je l'ai démontré dans ma thèse de Doctorat [Lop23b].

Jalon 7. Démontrer la terminaison des algorithmes de *Chase* sur des cas concrets à l'aide de résultats sur les systèmes de transition bien structurés.

4.3 Direction 2 : transductions

Dans cette partie, on s'intéresse plus précisément à la vérification des *pipelines* de manipulation de données structurées. Typiquement, l'entrée d'un tel pipeline est un fichier au format JSON, respectant un schéma prédéfini, et avec une profondeur maximale d'imbrication fixée à l'avance. Contraindre l'imbrication est une hypothèse raisonnable en pratique, et permet de garantir que la *sérialisation* et la *désérialisation* des données rentrent dans le cadre des fonctions polyrégulières.

Axe 4

Étudier la décidabilité des propriétés sémantiques des fonctions polyrégulières.

Une des questions ouvertes et fondamentales est de décider *l'équivalence de deux fonctions polyrégulières*, c'est-à-dire, étant donné deux fonctions f et g , décider si pour tout x , $f(x) = g(x)$. Cette question algorithmique est conjecturée décidable, et sa résolution permettrait de vérifier si deux pipelines sont équivalents, par exemple pour vérifier la correction d'une optimisation d'un processus de nettoyage de données. Les techniques usuelles pour décider l'équivalence de deux fonctions définies à l'aide d'automates ne sont pas applicables directement aux fonctions polyrégulières.¹⁵ Cependant, l'étude de ces transformations dans un cadre plus restreint (mots commutatifs) a déjà porté des fruits [BK24]. Par ailleurs, le problème est résolu lorsque l'alphabet de *sortie* est unaire, car dans ce cas, les fonctions polyrégulières forment un sous ensemble des séries rationnelles (qui seront définies plus bas) pour lequel l'équivalence est décidable [BR11].

Une première étape pour aborder ce problème est de se restreindre à des fonctions polyrégulières dont l'entrée est sur un alphabet unaire. Cette restriction fait passer de l'étude d'une *fonction* à l'étude d'une *suite* de mots $f(\varepsilon)$, $f(a)$, $f(aa)$, etc. J'ai déjà travaillé sur les simplifications apportées par une hypothèse *d'entrée commutative* sur les fonctions polyrégulières, c'est-à-dire,

13. $\forall x_1, x_2, x_3, x_4, x_5. \bigwedge_{1 \leq i \neq j \leq 5} x_i \neq x_j \wedge \neg \text{edge}(x_i, x_j) \Rightarrow \exists y. \bigwedge_{i=1}^5 \text{edge}(x_i, y)$.

14. Plus précisément, l'ensemble des états *plus petits* qu'un élément de $\text{Reach}(G)$. C'est-à-dire, sa *clôture par le bas*.

15. La méthode classique consistant au choix à exprimer les témoins de non-équivalence en utilisant une grammaire algébrique, ou en utilisant la décidabilité de l'équivalence des *automates polynomiaux*, un modèle d'automate quantitatif.

des fonctions polyrégulières f telles que $f(u) = f(v)$ dès lors que u et v ont la même image de Parikh.¹⁶ Toute fonction à entrée unaire est commutative par définition, mais la décomposition d'une fonction polyrégulière à entrée unaire en fonctions plus simples amène naturellement à considérer des fonctions ayant plusieurs paramètres d'entrée, d'où l'intérêt de cette hypothèse de travail.

Jalon 8. *Décider l'équivalence de deux fonctions polyrégulières commutatives.*

Une sous-classe des fonctions polyrégulières est d'une importance particulière : celle des fonctions polyrégulières où toutes les réaffectations ont la forme $\mathbf{b} = \text{True}$ et qui sont dites *apériodiques* [Boj18, Section 3]. Un problème fondamental de la théorie est de décider si une fonction polyrégulière f est apériodique.

Par exemple, la fonction `enrich_row_city` de l'Algorithme 1 est une fonction polyrégulière apériodique. Cette question est intéressante théoriquement car la notion d'apériodicité est liée aux théorèmes fondamentaux établissant l'équivalence entre les langages *sans étoile*, les automates *sans compteur*, et la *logique du premier ordre* sur les mots finis [Sch65b; PP86; MP71; voir Str18, pour une vision historique]. Dans le cadre du mouvement général des *langages* vers les *fonctions*, la décidabilité de ce problème augure en général une bonne compréhension théorique et algorithmique du modèle de calcul considéré.

De plus, la question de l'apériodicité est particulièrement intéressante dans ce projet de recherche car les fonctions polyrégulières apériodiques sont précisément celles qui peuvent s'exprimer en utilisant des *transductions* en logique du premier ordre FO [BKL19, Théorème 7]. L'intérêt de ces fonctions est donc double : d'une part, elles définissent des transformations qui sont compatibles avec la logique au sens de la définition donnée pour les systèmes centrés sur les données,¹⁷ et d'autre part, dans un objectif de vérification, elles peuvent s'exprimer dans un langage compatible avec les SMT solveurs (Satisfiability Modulo Theories) qui sont des outils efficaces pour résoudre des problèmes de décision en logique du premier ordre. L'idée est que des approches *incomplètes* pour la vérification utilisant des SMT solveurs pourraient s'avérer *utilisables* dans le cas où une approche *complète* utilisant la théorie des automates finis aurait une complexité trop élevée.

Jalon 9. *Décider si une fonction polyrégulière est apériodique.*

Ce problème est aussi ouvert dans le cas restreint des fonctions *linéaires régulières*. J'ai démontré avec COLCOMBET et DOUÉNEAU-TABOT [CDL23] que ce problème est décidable dans un cadre légèrement modifié. La clef de cette avancée majeure a été d'appliquer successivement une *restriction* du modèle de calcul, puis une *extension* du modèle de calcul : tout d'abord, en se restreignant à des fonctions polyrégulières dont la *sortie* est sur un alphabet d'une unique lettre, qui s'apparentent à des fonctions calculant des entiers naturels. Par la suite, en tirant profit de cette vision *quantitative* des valeurs de sorties pour développer un algorithme de décision dans le cas où les valeurs de sorties sont des *entiers relatifs*. Le passage de $\mathbb{N}$ à $\mathbb{Z}$ est un *pas de côté* qui permet de tirer profit des théories algébriques des semi-anneaux pour obtenir des résultats sur la décidabilité de problèmes algorithmiques comme l'apériodicité des fonctions polyrégulières. Malheureusement, le transfert de ces résultats à des fonctions à sortie dans $\mathbb{N}$ reste un problème ouvert qui nous amène à la conjecture suivante.

Jalon 10. *Démontrer qu'une fonction polyrégulière à sortie unaire est apériodique si et seulement si elle est apériodique vue comme une fonction polyrégulière à sortie dans $\mathbb{Z}$.*

Dans le cas des fonctions polyrégulières commutatives, cet objectif a déjà été atteint [Lop25]. Le passage au cas général pourrait se faire en utilisant les techniques de FO-approximants développées COLCOMBET, VAN GOOL et MORVAN [CVM22], en démontrant que cette approximation coïncide avec la fonction initiale dès que celle-ci est apériodique.

Une autre question d'intérêt, ouverte par ce *pas de côté*, est celle de l'apériodicité pour des modèles d'automates *quantitatifs*. En effet, les fonctions polyrégulières à sortie dans $\mathbb{Z}$ forment une sous-classe des *séries rationnelles*, qui sont des fonctions de Σ^* dans $\mathbb{Z}$ calculables par des automates finis pondérés par des entiers relatifs [Sch65b]. En réalité, ces fonctions sont exactement les

16. C'est-à-dire, lorsque u et v ont le même nombre d'occurrences de chaque lettre de l'alphabet d'entrée.

17. En effet, la pré-image d'un ensemble FO-définissable par une fonction polyrégulière apériodique est FO-définissable.

séries rationnelles dont la *croissance* est polynomiale.¹⁸ Définir une notion d'*apériodicité* pour les séries rationnelles qui coïncide avec la notion d'apériodicité pour les fonctions polyrégulières à sortie dans $\mathbb{Z}$ n'avait pas été envisagé jusqu'à présent.

À toute série rationnelle, on peut associer une *représentation linéaire minimale* sous la forme d'un morphisme $\mu: \Sigma^* \rightarrow \mathcal{M}_{n \times n}(\mathbb{Z})$ qui associe à chaque lettre de l'alphabet une matrice entière. Je propose de décider qu'une série rationnelle est apériodique si et seulement si les *valeurs propres* de $\mu(w)$ sont toutes réelles positives.¹⁹ Cette définition coïncide avec celle obtenue pour les fonctions polyrégulières à sortie dans $\mathbb{Z}$, qui ont comme caractéristique supplémentaire d'avoir des valeurs propres de module²⁰ 0 ou 1 [CDL23].

Jalon 11. *Décider si une série rationnelle est apériodique.*

Je pense que cette direction de recherche est particulièrement intéressante car elle permet de relier des questions *quantitatives* (valeurs propres) à des questions traditionnellement *qualitatives* (apériodicité). Par ailleurs, les résultats récents sur la décidabilité de propriétés non-triviales des séries rationnelles comme la *déterminisabilité* ou la *désambiguïsité* à l'aide des techniques de géométrie algébrique (clôtures de Zariski) semblent fournir une piste pour les approcher [BS23; PS24].

Finalement, cette question de l'apériodicité rejoindra aussi une question plus générale sur les séries rationnelles, analogue à celle sur les fonctions polyrégulières à valeurs dans $\mathbb{Z}$: quelles sont ces fonctions qui sont en réalité définissables dans $\mathbb{N}$? Et est-ce que les résultats *quantitatifs* (sur $\mathbb{Z}$) peuvent être transférés à des résultats *qualitatifs* (sur $\mathbb{N}$) ?

Axe 5

Implémenter des outils de vérification de fonctions polyrégulières, et l'appliquer à la vérification de pipelines ETL.

L'objectif de cet axe est de tirer profit des résultats théoriques existants pour construire des prototypes d'outils de vérification de transformations de données. Concrètement, si l'équivalence entre deux fonctions polyrégulières n'est pas un problème aujourd'hui connu comme décidable, il est en revanche possible de vérifier des *spécifications* sur ces fonctions sous la forme de triplets de Hoare $\{P\} \text{ code } \{Q\}$, où P et Q sont des formules de la logique MSO sur les mots. Comme expliqué dans l'introduction de ce projet de recherche, aucune implémentation n'existe à ce jour.

Jalon 12. *Construire un prototype de programme vérifiant des fonctions polyrégulières écrites en Python.*

Une grande partie du travail consistera à trouver un équilibre entre l'expressivité du sous-ensemble de Python considéré et la facilité à le transformer en une fonction polyrégulière représentée par une transduction MSO. L'intérêt de considérer Python comme langage cible est qu'une grande partie des transformations de données sont déjà écrites dans ce langage, et que sa syntaxe est adaptée à la description de fonctions polyrégulières. En effet, les fonctions polyrégulières peuvent être comprises comme des fonctions manipulant des listes imbriquées en utilisant uniquement des variables booléennes et la syntaxe de boucles sur les listes `for (i, v) in enumerate(1): ...` de Python. Ultimement, l'intégration dans une suite de tests automatiques permettrait d'agir en complément des outils existants (tests unitaires, tests de propriétés, etc.) pour garantir une qualité de code élevée.

Ainsi, une première étape avant de construire un outil vérifiant un sous ensemble de Python est de définir un langage de programmation sur les types `list`, `dict`, et `str`, possédant des fonctions, des boucles, et des conditions, et dont la syntaxe soit suffisamment proche de Python pour être utilisable en attendant de construire un outil de vérification qui s'intégrera dans un environnement de développement.

Jalon 13. *Formaliser un langage de programmation haut-niveau décrivant des fonctions polyrégulières.*

18. Une fonction est à croissance polynomiale s'il existe un polynôme P tel que pour tout mot w , $|f(w)| \leq P(|w|)$.

19. Une valeur propre d'une matrice M est un nombre λ tel qu'il existe un vecteur non nul v tel que $Mv = \lambda v$. Une matrice à coefficients dans $\mathbb{Z}$ possède des valeurs propres dans $\mathbb{C}$, et une valeur propre est dite *réelle positive* lorsque sa partie imaginaire est nulle et sa partie réelle positive.

20. Le module d'un nombre complexe est sa distance à l'origine dans le plan complexe.

En collaboration avec Rafał STEFAŃSKI de l'Université de Varsovie, je travaille sur la vérification de transformations de chaînes de caractères dans un langage très similaire à Python [LS24], en se restreignant aux fonctions polyrégulières apériodiques. Les premiers résultats semblent indiquer que les obligations de preuve générées par les triplets de Hoare peuvent être résolues efficacement par des méthodes de théorie des automates et des SMT solveurs pour la logique du premier ordre (comme CVC5 et Z3). Par ailleurs, la compilation de ces fonctions vers la logique du premier ordre possède un autre intérêt pratique : on sait théoriquement minimiser le nombre de variables utilisées dans une telle présentation, et les transductions FO sont extrêmement parallélisables. Ainsi, une autre piste d'investigation est la compilation de ces fonctions vers des modèles *parallèles* et *efficaces* pour le traitement massif de données.

Jalon 14. *Évaluer la possibilité d'évaluation efficace de fonctions polyrégulières sur des données massives.*

4.4 Intégration

Je souhaite intégrer l'équipe LX du département M2F (Méthodes et Modèles Formels), dont les centres d'intérêt sont en parfait alignement avec les miens. En effet, mes travaux de recherche passés et mon projet de recherche sont connectés aux thématiques de l'équipe.

Pour ce qui est de la théorie des transducteurs et des automates où je pourrais collaborer avec Thomas Place, Marc Zeitoun en lien avec la partie de mon projet concernant les notions d'apériodicité pour des *transducteurs*. Par ailleurs, l'équipe LX possède une expertise en transducteurs et automates pondérés par la présence de Sylvain Lombardy et Vincent Penelle.

Une seconde partie de mon projet de recherche concerne les problèmes de beaux préordres et systèmes de transitions dont l'espace des états est une classe de structures relationnelles finies. Dans ce cadre, je pense que je pourrais collaborer avec Jérôme Leroux, qui est un expert en théorie des beaux préordres. Dans ce même cadre, je pourrais également collaborer avec les membres de l'équipe RATIO en lien avec la partie de mon projet concernant les potentielles interactions avec les algorithmes de complétion de bases de données.

Par ailleurs, je collabore déjà avec un membre du département M2F, à savoir Arka Ghosh, sur un projet visant à généraliser les résultats classiques de la théorie des polynômes (théorème de Hilbert, calcul de bases de Gröbner) à des polynômes ayant une infinité de variables, en utilisant des outils de la théorie des « orbit finite sets », en partie développée à Varsovie.

4.4.1 Collaborations nationales et internationales

Au niveau national, j'entretiens des liens avec l'équipe Automates de l'IRIF (Thomas COLCOMBET, Sylvain SCHMITZ), le LMF (Jean GOUBAULT-LARRECQ), l'équipe Synthèse à Bordeaux (Arka GHOSH), et l'équipe MOVE à Marseille (Nathan LHOTE). Par ailleurs, il serait intéressant pour moi de nouer des collaborations avec l'équipe VALDA de l'INRIA Paris dirigée par Pierre SENELLART, l'équipe LINKS de l'INRIA Lille dirigée par Sylvain SALVATI concernant la partie de mon projet faisant le lien entre l'évaluation de requête sur des bases de données incomplètes et les systèmes centrés sur les données. Au niveau international, je souhaite pérenniser mes liens avec l'Université de Varsovie (Sławomir LASOTA, Mikołaj BOJAŃCZYK, Rafał STEFAŃSKI, Szymon TORUŃCZYK) et renforcer mes collaborations avec le « Max Planck Institute for Software Systems » (MPI-SWS), notamment en continuant ma collaboration avec Lia SCHÜTZE.

Références

Références personnelles publiées (citées)

- [Lop21] Aliaume LOPEZ. « Preservation Theorems Through the Lens of Topology ». In : *Proceedings of CSL21*. T. 183. 2021, 32 :1-32 :17. DOI : [10.4230/LIPIcs.CSL.2021.32](https://doi.org/10.4230/LIPIcs.CSL.2021.32).
- [Lop22] Aliaume LOPEZ. « When Locality Meets Preservation ». In : *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*. New York, NY, USA, août 2022, p. 1-14. DOI : [10.1145/3531130.3532498](https://doi.org/10.1145/3531130.3532498). URL : <https://doi.org/10.1145/3531130.3532498>.
- [CDL23] T. COLCOMBET, G. DOUENEAU-TABOT et A. LOPEZ. « $\mathbb{Z}$ -polyregular functions ». In : *2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. Los Alamitos, CA, USA, juin 2023, p. 1-13. DOI : [10.1109/LICS56636.2023.10175685](https://doi.ieeecomputersociety.org/10.1109/LICS56636.2023.10175685). URL : <https://doi.ieeecomputersociety.org/10.1109/LICS56636.2023.10175685>.
- [GHL23] Jean GOUBAULT-LARRECQ, Simon HALFON et Aliaume LOPEZ. « Infinitary Noetherian Constructions II. Transfinite Words and the Regular Subword Topology ». In : *Colloquium Mathematicum* 173.2 (2023), p. 227-250. DOI : [10.4064/cm8793-3-2023](https://hal.science/hal-04283731). URL : <https://hal.science/hal-04283731>.
- [Lop23a] Aliaume LOPEZ. « Fixed Points and Noetherian Topologies ». In : *Foundations of Software Science and Computation Structures*. T. 13992. 2023, p. 456-476. DOI : [10.1007/978-3-031-30829-1_22](https://link.springer.com/chapter/10.1007/978-3-031-30829-1_22). URL : https://link.springer.com/chapter/10.1007/978-3-031-30829-1_22.
- [Lop23b] Aliaume LOPEZ. « Théorèmes de préservation pour la logique au premier ordre : localité, topologie et constructions limites. » Thèse de doct. 2023. URL : <https://www.irif.fr/~alopez/ressources/these/defense/2023-09-12-defense-manuscript.pdf>.

Références personnelles publiées (non citées)

- [GJL17] Dan R. GHICA, Achim JUNG et Aliaume LOPEZ. « Diagrammatic Semantics for Digital Circuits ». In : *26th EACSL Annual Conference on Computer Science Logic (CSL 2017)*. T. 82. Dagstuhl, Germany, 2017, 24 :1-24 :16. DOI : [10.4230/LIPIcs.CSL.2017.24](http://drops.dagstuhl.de/opus/volltexte/2017/7671). URL : <http://drops.dagstuhl.de/opus/volltexte/2017/7671>.
- [GL17] Dan R. GHICA et Aliaume LOPEZ. « A structural and nominal syntax for diagrams ». In : *14th International Conference on Quantum Physics and Logic, QPL 2017, Nijmegen, The Netherlands, 3-7 July 2017*. T. 266. 2017, p. 71-83. DOI : [10.4204/EPTCS.266.4](https://doi.org/10.4204/EPTCS.266.4). URL : <https://doi.org/10.4204/EPTCS.266.4>.
- [LS18] Aliaume LOPEZ et Alex SIMPSON. « Basic Operational Preorders for Algebraic Effects in General, and for Combined Probability and Nondeterminism in Particular ». In : *27th EACSL Annual Conference on Computer Science Logic (CSL 2018)*. T. 119. Dagstuhl, Germany, 2018, 29 :1-29 :17. DOI : [10.4230/LIPIcs.CSL.2018.29](http://drops.dagstuhl.de/opus/volltexte/2018/9696). URL : <http://drops.dagstuhl.de/opus/volltexte/2018/9696>.

Références personnelles non-publiées (citées)

- [Abr+23] Sergio ABRIOLA, Simon HALFON, Aliaume LOPEZ, Sylvain SCHMITZ, Philippe SCHNOEBELEN et Isa VIALARD. *Measuring well quasi-ordered finitary powersets*. To appear in Math. Sci. Comput. 2023. arXiv : [2312.14587](https://arxiv.org/abs/2312.14587) [cs.LO].
- [LLS24] Nathan LHOUE, Aliaume LOPEZ et Lia SCHÜTZE. *Well-quasi-orderings on word languages*. 2024. arXiv : [2501.07428v1](https://github.com/AliaumeL/word-ordering-wqo) [cs.FL]. URL : <https://github.com/AliaumeL/word-ordering-wqo>.
- [Lop24a] Aliaume LOPEZ. *$\mathbb{N}$ -polyregular functions arise from well-quasi-orderings*. 2024. DOI : [10.48550/ARXIV.2409.07882](https://arxiv.org/abs/2409.07882v1). arXiv : [2409.07882v1](https://arxiv.org/abs/2409.07882v1) [cs.LO]. URL : <https://arxiv.org/abs/2409.07882v1>.
- [Lop24b] Aliaume LOPEZ. *Labelled Well Quasi Ordered Classes of Bounded Linear Clique-Width*. 2024. arXiv : [2405.10894](https://arxiv.org/abs/2405.10894) [cs.LO]. URL : <https://arxiv.org/abs/2405.10894>.
- [LS24] Aliaume LOPEZ et Rafał STEFAŃSKI. *First-Order Polyregular Functions and Model Checking*. Version 0.0.1. 3 sept. 2024. URL : <https://github.com/AliaumeL/polyregular-model-checking>.

Autres références

- [Hig52] Graham HIGMAN. « Ordering by divisibility in abstract algebras ». In : *Proceedings of the London Mathematical Society* 3 (1952), p. 326-336. DOI : [10.1112/plms/s3-2.1.326](https://doi.org/10.1112/plms/s3-2.1.326).
- [Rad54] R. RADO. « Partial well-ordering of sets of vectors ». In : *Mathematika* 1.2 (1954), p. 89-95. DOI : [10.1112/S0025579300000565](https://doi.org/10.1112/S0025579300000565).
- [Tai59] William W. TAIT. « A counterexample to a conjecture of Scott and Suppes ». In : *Journal of Symbolic Logic* 24 (1959), p. 15-16. DOI : [10.2307/2964569](https://doi.org/10.2307/2964569).
- [Sch62] Marcel Paul SCHÜTZENBERGER. « Finite Counting Automata ». In : *Information and control* 5.2 (1962), p. 91-107.
- [Nas65] Crispin relax St. John Alvah NASH-WILLIAMS. « On well-quasi-ordering transfinite sequences ». In : *Mathematical Proceedings of the Cambridge Philosophical Society* 61 (1965), p. 33-39.
- [Sch65b] M. P. SCHÜTZENBERGER. « On finite monoids having only trivial subgroups ». en. In : *Information and Control* 8.2 (avr. 1965), p. 190-194. DOI : [10.1016/S0019-9958\(65\)90108-7](https://doi.org/10.1016/S0019-9958(65)90108-7).
- [Sco67] Dana SCOTT. « Some definitional suggestions for automata theory ». In : *Journal of Computer and System Sciences* 1.2 (août 1967), p. 187-212. DOI : [10.1016/S0022-0000\(67\)80014-X](https://doi.org/10.1016/S0022-0000(67)80014-X). URL : <https://www.sciencedirect.com/science/article/pii/S002200006780014X>.
- [KM69] Richard M. KARP et Raymond E. MILLER. « Parallel program schemata ». In : *Journal of Computer and System Sciences* 3.2 (1969), p. 147-195. DOI : [https://doi.org/10.1016/S0022-0000\(69\)80011-5](https://doi.org/10.1016/S0022-0000(69)80011-5). URL : <https://www.sciencedirect.com/science/article/pii/S0022000069800115>.
- [Cod70] Edgar F. CODD. « A relational model of data for large shared data banks ». In : *Commun. ACM* 13.6 (1970), p. 377-387. DOI : [10.1145/362384.362685](https://doi.org/10.1145/362384.362685).
- [MP71] Robert McNAUGHTON et Seymour A PAPERT. *Counter-Free Automata*. 1971.
- [Kru72] Joseph B. KRUSKAL. « The theory of well-quasi-ordering : A frequently discovered concept ». In : *Journal of Combinatorial Theory, Series A* 13 (1972), p. 297-305. DOI : [10.1016/0097-3165\(72\)90063-5](https://doi.org/10.1016/0097-3165(72)90063-5).
- [Pou72] Maurice POUZET. « Un bel ordre d'abritement et ses rapports avec les bornes d'une multirelation ». In : *CR Acad. Sci. Paris Sér. AB* 274 (1972), A1677-A1680.
- [DP77] Dick H. J. DE JONGH et Rohit PARIKH. « Well-partial orderings and hierarchies ». In : *Indag. Math.* 39.3 (1977), p. 195-207. DOI : [10.1016/1385-7258\(77\)90067-1](https://doi.org/10.1016/1385-7258(77)90067-1).
- [Kar77] Juhani KARHUMÄKI. « Remarks on commutative n-rational series ». In : *Theoretical Computer Science* 5.2 (1977), p. 211-217. DOI : [https://doi.org/10.1016/0304-3975\(77\)90008-1](https://doi.org/10.1016/0304-3975(77)90008-1). URL : <https://www.sciencedirect.com/science/article/pii/0304397577900081>.
- [Reu80] Christophe REUTENAUER. « Séries formelles et algèbres syntactiques ». In : *Journal of Algebra* 66.2 (1980), p. 448-483. DOI : [https://doi.org/10.1016/0021-8693\(80\)90097-6](https://doi.org/10.1016/0021-8693(80)90097-6). URL : <https://www.sciencedirect.com/science/article/pii/0021869380900976>.
- [Sch81] D. SCHMIDT. « The relation between the height of a well-founded partial ordering and the order types of its chains and antichains ». In : *Journal of Combinatorial Theory, Series B* 31.2 (1981), p. 183-189. DOI : [10.1016/S0095-8956\(81\)80023-8](https://doi.org/10.1016/S0095-8956(81)80023-8).
- [Gai82] Haim GAIFMAN. « On local and non-local properties ». In : *Proceedings of the Herbrand Symposium*. T. 107. 1982, p. 105-135. DOI : [10.1016/S0049-237X\(08\)71879-2](https://doi.org/10.1016/S0049-237X(08)71879-2).
- [Mil85] Eric C MILNER. « Basic wqo-and bqo-theory ». In : *Graphs and order*. 1985, p. 487-502.
- [PP86] Dominique PERRIN et Jean-Eric PIN. « First-order logic and star-free sets ». In : *Journal of Computer and System Sciences* 32.3 (1986), p. 393-406.
- [AG87] Miklós AJTAI et Yuri GUREVICH. « Monotone versus positive ». In : *Journal of the ACM* 34 (1987), p. 1004-1015. DOI : [10.1145/31846.31852](https://doi.org/10.1145/31846.31852).

- [CK90] Chen Chung CHANG et H. Jerome KEISLER. *Model Theory*. T. 73. 1990.
- [KT90] I. KRÍŽ et R. THOMAS. « Ordinal Types in Ramsey Theory and Well-Partial-Ordering Theory ». In : *Mathematics of Ramsey Theory*. T. 5. 1990, p. 57-95. DOI : [10.1007/978-3-642-72905-8_7](https://doi.org/10.1007/978-3-642-72905-8_7).
- [Sim90a] Imre SIMON. « Factorization Forests of Finite Height ». In : *Theor. Comput. Sci.* 72.1 (1990), p. 65-94. DOI : [10.1016/0304-3975\(90\)90047-L](https://doi.org/10.1016/0304-3975(90)90047-L).
- [Sim90b] Imre SIMON. « Factorization forests of finite height ». In : *Theoretical Computer Science* 72.1 (1990), p. 65-94. DOI : [10.1016/0304-3975\(90\)90047-L](https://doi.org/10.1016/0304-3975(90)90047-L). URL : <https://www.sciencedirect.com/science/article/pii/030439759090047L>.
- [Din92] Guoli DING. « Subgraphs and well-quasi-ordering ». In : *Journal of Graph Theory* 16 (1992), p. 489-502. DOI : [10.1002/jgt.3190160509](https://doi.org/10.1002/jgt.3190160509).
- [AG94] Miklós AJTAI et Yuri GUREVICH. « Datalog vs first-order logic ». In : *Journal of Computer and System Sciences* 49 (1994), p. 562-588. DOI : [10.1016/S0022-0000\(05\)80071-6](https://doi.org/10.1016/S0022-0000(05)80071-6).
- [AHV95] Serge ABITEBOUL, Richard HULL et Victor VIANU. *Foundations of Databases*. 1995.
- [Sto95] Alexei P. STOLBOUSHKIN. « Finitely monotone properties ». In : (1995), p. 324-330. DOI : [10.1109/LICS.1995.523267](https://doi.org/10.1109/LICS.1995.523267).
- [Abd+96] Parosh Aziz ABDULLA, Karlis ČERĀNS, Bengt Jonsson TSAY et YIH-KUEN. « General decidability theorems for infinite-state systems ». In : *Proceedings of LICS'96*. 1996, p. 313-321. DOI : [10.1109/LICS.1996.561359](https://doi.org/10.1109/LICS.1996.561359).
- [Gal97] Jean H. GALLIER. « Ann. Pure Appl. Logic : Erratum to “What’s so special about Kruskal’s Theorem and the ordinal Γ_0 ? A survey of some results in proof theory” [53 (1991) 199–260] ». In : *Annals of Pure and Applied Logic* 89 (1997), p. 275. DOI : [10.1016/S0168-0072\(97\)00043-2](https://doi.org/10.1016/S0168-0072(97)00043-2).
- [AJ98] Parosh Aziz ABDULLA et Bengt JONSSON. « Verifying networks of timed processes ». In : *Proceedings of TACAS'98*. T. 1384. 1998, p. 298-312. DOI : [10.1007/BFb0054179](https://doi.org/10.1007/BFb0054179).
- [AB99] U. ABRAHAM et R. BONNET. « Hausdorff’s theorem for posets that satisfy the finite antichain property ». In : *Fund. Math.* 159.1 (1999), p. 51-69. DOI : [10.4064/fm-159-1-51-69](https://doi.org/10.4064/fm-159-1-51-69).
- [Jan99] Petr JANČAR. « A note on well quasi-orderings for powersets ». In : *Information Processing Letters* 72 (1999), p. 155-160. DOI : [10.1016/S0020-0190\(99\)00149-0](https://doi.org/10.1016/S0020-0190(99)00149-0).
- [FS01] Alain FINKEL et Philippe SCHNOEBELEN. « Well-structured transition systems everywhere! ». In : *Theoretical Computer Science* 256 (2001), p. 63-92. DOI : [10.1016/S0304-3975\(00\)00102-X](https://doi.org/10.1016/S0304-3975(00)00102-X). URL : <https://linkinghub.elsevier.com/retrieve/pii/S030439750000102X>.
- [Has02] Ryu HASEGAWA. « Two applications of analytic functors ». In : *Theoretical Computer Science* 272 (2002), p. 113-175. DOI : [10.1016/S0304-3975\(00\)00349-2](https://doi.org/10.1016/S0304-3975(00)00349-2). URL : <https://www.sciencedirect.com/science/article/pii/S0304397500003492>.
- [DT03] Nachum DERSHOWITZ et Iddo TZAMERET. « Gap Embedding for Well-Quasi-Orderings ». In : *Electronic Notes in Theoretical Computer Science* 84 (2003). WoLLIC’2003, 10th Workshop on Logic, Language, Information and Computation, p. 80-90. DOI : [10.1016/S1571-0661\(04\)80846-6](https://doi.org/10.1016/S1571-0661(04)80846-6). URL : <https://www.sciencedirect.com/science/article/pii/S1571066104808466>.
- [RS04] Neil ROBERTSON et P.D. SEYMOUR. « Graph Minors. XX. Wagner’s conjecture ». In : *Journal of Combinatorial Theory, Series B* 92.2 (2004). Special Issue Dedicated to Professor W.T. Tutte, p. 325-357. DOI : <https://doi.org/10.1016/j.jctb.2004.08.001>. URL : <https://www.sciencedirect.com/science/article/pii/S0095895604000784>.
- [ADK06] Albert ATSERIAS, Anuj DAWAR et Phokion G. KOLAITIS. « On preservation under homomorphisms and unions of conjunctive queries ». In : *Journal of the ACM* 53 (2006), p. 208-237. DOI : [10.1145/1131342.1131344](https://doi.org/10.1145/1131342.1131344).
- [Col07] Thomas COLCOMBET. « A Combinatorial Theorem for Trees ». In : *Automata, Languages and Programming*. Berlin, Heidelberg, 2007, p. 901-912.

- [Gou07] Jean GOUBAULT-LARRECQ. « On Noetherian spaces ». In : *Proceedings of LICS'07*. 2007, p. 453-462. DOI : [10.1109/LICS.2007.34](https://doi.org/10.1109/LICS.2007.34).
- [ADGo8] Albert ATSERIAS, Anuj DAWAR et Martin GROHE. « Preservation under extensions on well-behaved finite structures ». In : *SIAM Journal on Computing* 38 (2008), p. 1364-1381. DOI : [10.1137/060658709](https://doi.org/10.1137/060658709).
- [DNRo8a] Alin DEUTSCH, Alan NASH et Jeffrey B. REMMEL. « The chase revisited ». In : *Proceedings of PODS'08*. 2008, p. 149-158. DOI : [10.1145/1376916.1376938](https://doi.org/10.1145/1376916.1376938).
- [Roso8] Benjamin ROSSMAN. « Homomorphism preservation theorems ». In : *Journal of the ACM* 55 (2008), 15:1-15:53. DOI : [10.1145/1379759.1379763](https://doi.org/10.1145/1379759.1379763).
- [Wei09] A. WEIERMANN. « A Computation of the Maximal Order Type of the Term Ordering on Finite Multisets ». In : *Proc. 5th Conf. Computability in Europe (CiE 2009), Heidelberg, Germany, July 2009*. T. 5635. 2009, p. 488-498. DOI : [10.1007/978-3-642-03073-4_50](https://doi.org/10.1007/978-3-642-03073-4_50).
- [DRT10] Jean DALIGAUT, Michael RAO et Stéphan THOMASSÉ. « Well-Quasi-Order of Relabel Functions ». In : *Order* 27 (2010), p. 301-315. DOI : [10.1007/s11083-010-9174-0](https://doi.org/10.1007/s11083-010-9174-0). URL : <https://doi.org/10.1007/s11083-010-9174-0>.
- [Daw10] Anuj DAWAR. « Homomorphism preservation on quasi-wide classes ». In : *Journal of Computer and System Sciences* 76 (2010), p. 324-332. DOI : [10.1016/j.jcss.2009.10.005](https://doi.org/10.1016/j.jcss.2009.10.005).
- [Gou10] Jean GOUBAULT-LARRECQ. « Noetherian Spaces in Verification ». In : *Automata, Languages and Programming, 37th International Colloquium, ICALP 2010, Bordeaux, France, July 6-10, 2010, Proceedings, Part II*. T. 6199. 2010, p. 2-21. DOI : [10.1007/978-3-642-14162-1_2](https://doi.org/10.1007/978-3-642-14162-1_2). URL : https://doi.org/10.1007/978-3-642-14162-1_2.
- [BR11] Jean BERSTEL et Christophe REUTENAUER. *Noncommutative rational series with applications*. T. 137. 2011.
- [Fig+11b] D. FIGUEIRA, S. FIGUEIRA, S. SCHMITZ et Ph. SCHNOEBELEN. « Ackermannian and Primitive-Recursive Bounds with Dickson's Lemma ». In : *Proc. 26th IEEE Symp. Logic in Computer Science (LICS 2011), Toronto, Canada, June 2011*. 2011, p. 269-278. DOI : [10.1109/LICS.2011.39](https://doi.org/10.1109/LICS.2011.39).
- [CGL12] Andrea CALÌ, Georg GOTTLÖB et Thomas LUKASIEWICZ. « A general Datalog-based framework for tractable query answering over ontologies ». In : *Journal of Web Semantics* 14 (2012). Special Issue on Dealing with the Messiness of the Web of Data, p. 57-83. DOI : <https://doi.org/10.1016/j.websem.2012.03.001>. URL : <https://www.sciencedirect.com/science/article/pii/S1570826812000388>.
- [CE12] Bruno COURCELLE et Joost ENGELFRIET. *Graph structure and monadic second-order logic : a language-theoretic approach*. T. 138. 2012.
- [Dem+12] Stéphane DEMERI, Alain FINKEL, Jean GOUBAULT-LARRECQ, Sylvain SCHMITZ et Philippe SCHNOEBELEN. « Algorithmic Aspects of WQO Theory (MPRI course) ». Master Parisien de Recherche en Informatique, 2012. URL : <https://cel.archives-ouvertes.fr/cel-00727025>.
- [Bon+13] R. BONNET, A. FINKEL, S. HADDAD et F. ROSA-VELARDO. « Ordinal theory for expressiveness of well-structured transition systems ». In : *Information and Computation* 224 (2013), p. 1-22. DOI : [10.1016/j.ic.2012.11.003](https://doi.org/10.1016/j.ic.2012.11.003).
- [KR13] Stephan KREUTZER et Cristian RIVEROS. « Quantitative Monadic Second-Order Logic ». In : *2013 28th Annual ACM/IEEE Symposium on Logic in Computer Science*. 2013, p. 113-122. DOI : [10.1109/LICS.2013.16](https://doi.org/10.1109/LICS.2013.16).
- [SAC14] Abhisekh SANKARAN, Bharat ADSUL et Supratik CHAKRABORTY. « A generalization of the Łoś-Tarski preservation theorem over classes of finite structures ». In : *International Symposium on Mathematical Foundations of Computer Science*. 2014, p. 474-485. DOI : [10.1007/978-3-662-44522-8_40](https://doi.org/10.1007/978-3-662-44522-8_40).
- [AFS15] S. ABRIOLA, S. FIGUEIRA et G. SENNO. « Linearizing well quasi-orders and bounding the length of bad sequences ». In : *Theor. Comp. Sci.* 603 (2015), p. 3-22. DOI : [10.1016/j.tcs.2015.07.012](https://doi.org/10.1016/j.tcs.2015.07.012).
- [VRW15] J. VAN DER MEEREN, M. RATHJEN et A. WEIERMANN. « Well-partial-orderings and the big Veblen number ». In : *Archive for Mathematical Logic* 54.1-2 (2015), p. 193-230. DOI : [10.1007/s00153-014-0408-5](https://doi.org/10.1007/s00153-014-0408-5).

- [SAC16] Abhisekh SANKARAN, Bharat ADSUL et Supratik CHAKRABORTY. « A generalization of the Łoś–Tarski preservation theorem ». In : *Annals of Pure and Applied Logic* 167 (2016), p. 189–210.
- [Abd+17] Parosh ABDULLA, Alain FINKEL, Bengt JOHNSSON et Philippe SCHNOEBELE. *CAV Award*. 2017. URL : <https://i-cav.org/cav-award/>.
- [Sch17] Sylvain SCHMITZ. « Algorithmic Complexity of Well Quasi-Orders ». Habilitation à diriger des recherches. École normale supérieure Paris-Saclay, nov. 2017. URL : <https://theses.hal.science/tel-01663266>.
- [SF17] Luc SEGOUFIN et Diego FIGUEIRA. « Bottom-up automata on data trees and vertical XPath ». In : *Logical Methods in Computer Science* 13 (2017). DOI : [10.23638/LMCS-13\(4:5\)2017](https://doi.org/10.23638/LMCS-13(4:5)2017).
- [Boj18] Mikołaj BOJAŃCZYK. *Polyregular Functions*. 2018. DOI : [10.48550/ARXIV.1810.08760](https://arxiv.org/abs/1810.08760). URL : <https://arxiv.org/abs/1810.08760>.
- [Hal18] Simon HALFON. « On Effective Representations of Well Quasi-Orderings ». Theses. Université Paris Saclay (COMUE), juin 2018. URL : <https://theses.hal.science/tel-01945232>.
- [Str18] Howard STRAUBING. « First-order logic and aperiodic languages : a revisionist history ». In : *ACM SIGLOG News* 5,3 (juill. 2018), p. 4–20. DOI : [10.1145/3242953.3242956](https://doi.org/10.1145/3242953.3242956). URL : <https://doi.org/10.1145/3242953.3242956>.
- [BKL19] Mikołaj BOJAŃCZYK, Sandra KIEFER et Nathan LHOE. « String-to-String Interpretations With Polynomial-Size Output ». In : *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. T. 132. Dagstuhl, Germany, 2019, 106 :1–106 :14. DOI : [10.4230/LIPIcs.ICALP.2019.106](https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2019.106). URL : <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2019.106>.
- [BC19] Simone BOVA et Hubie CHEN. « How Many Variables are Needed to Express an Existential Positive Query? ». In : t. 63. 2019, p. 1573–1594. DOI : [10.1007/s00224-018-9884-z](https://doi.org/10.1007/s00224-018-9884-z). URL : <https://doi.org/10.1007/s00224-018-9884-z>.
- [DST19] Max DICKMANN, Niels SCHWARTZ et Marcus TRESSL. *Spectral Spaces*. T. 35. 2019. DOI : [10.1017/9781316543870.003](https://doi.org/10.1017/9781316543870.003).
- [DG19] Manfred DROSTE et Paul GASTIN. « Aperiodic Weighted Automata and Weighted First-Order Logic ». In : *44th International Symposium on Mathematical Foundations of Computer Science, MFCS 2019*. T. 138. 2019.
- [FG19] Alain FINKEL et Ekanshdeep GUPTA. « The Well Structured Problem for Presburger Counter Machines ». In : *39th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2019)*. T. 150. Dagstuhl, Germany, 2019, 41 :1–41 :15. DOI : [10.4230/LIPIcs.FSTTCS.2019.41](https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.FSTTCS.2019.41). URL : <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.FSTTCS.2019.41>.
- [MP19] Anca MUSCHOLL et Gabriele PUPPIS. « The Many Facets of String Transducers ». In : *36th International Symposium on Theoretical Aspects of Computer Science (STACS 2019)*. T. 126. Dagstuhl, Germany, 2019, 2 :1–2 :21. DOI : [10.4230/LIPIcs.STACS.2019.2](https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.STACS.2019.2). URL : <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.STACS.2019.2>.
- [Sch19a] S. SCHMITZ. « The Parametric Complexity of Lossy Counter Machines ». In : *Proc. 46th Int. Coll. Automata, Languages, and Programming (ICALP 2019), Patras, Greece, July 2019*. T. 132. 2019, 129 :1–129 :15. DOI : [10.4230/LIPIcs.ICALP.2019.129](https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2019.129).
- [Sch19b] Sylvain SCHMITZ. « The Parametric Complexity of Lossy Counter Machines ». In : *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. T. 132. Dagstuhl, Germany, 2019, 129 :1–129 :15. DOI : [10.4230/LIPIcs.ICALP.2019.129](https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2019.129). URL : <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2019.129>.
- [Bal20] A. R. BALASUBRAMANIAN. « Complexity of Controlled Bad Sequences over Finite Sets of $\mathbb{N}^d$ ». In : *Proc. 35th ACM/IEEE Symp. Logic in Computer Science (LICS 2020), Saarbrücken, Germany, July 2020*. 2020, p. 130–140. DOI : [10.1145/3373718.3394753](https://doi.org/10.1145/3373718.3394753).
- [DSS20a] M. DŽAMONJA, S. SCHMITZ et Ph. SCHNOEBELE. « On Ordinal Invariants in Well Quasi Orders and Finite Antichain Orders ». In : *Well Quasi-Orders in Computation, Logic, Language and Reasoning*. T. 53. 2020. Chap. 2, p. 29–54. DOI : [10.1007/978-3-030-30229-0_2](https://doi.org/10.1007/978-3-030-30229-0_2).

- [DSS20b] Mirna DŽAMONJA, Sylvain SCHMITZ et Philippe SCHNOEBELEN. « On Ordinal Invariants in Well Quasi Orders and Finite Antichain Orders ». In : *Well-Quasi Orders in Computation, Logic, Language and Reasoning : A Unifying Concept of Proof Theory, Automata Theory, Formal Languages and Descriptive Set Theory*. Cham, 2020, p. 29-54. DOI : [10.1007/978-3-030-30229-0_2](https://doi.org/10.1007/978-3-030-30229-0_2). URL : https://doi.org/10.1007/978-3-030-30229-0_2.
- [FG20] Alain FINKEL et Jean GOUBAULT-LARRECQ. « Forward analysis for WSTS, part I : completions ». In : *Mathematical Structures in Computer Science* 30 (2020), p. 752-832. DOI : [10.1017/S0960129520000195](https://doi.org/10.1017/S0960129520000195).
- [Sch20] D. SCHMIDT. « Well-Partial Orderings and their Maximal Order Types ». In : *Well Quasi-Orders in Computation, Logic, Language and Reasoning*. T. 53. 2020. Chap. 12, p. 351-391. DOI : [10.1007/978-3-030-30229-0_13](https://doi.org/10.1007/978-3-030-30229-0_13).
- [CF21] Yijia CHEN et Jörg FLUM. « Forbidden Induced Subgraphs and the Łoś-Tarski Theorem ». In : *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. 2021, p. 1-13. DOI : [10.1109/LICS52264.2021.9470742](https://doi.org/10.1109/LICS52264.2021.9470742).
- [DS21] Anuj DAWAR et Abhisekh SANKARAN. « Extension Preservation in the Finite and Prefix Classes of First Order Logic ». In : *29th EACSL Annual Conference on Computer Science Logic (CSL 2021)*. T. 183. 2021, 18 :1-18 :13. DOI : [10.4230/LIPIcs.CSL.2021.18](https://doi.org/10.4230/LIPIcs.CSL.2021.18). URL : <https://drops.dagstuhl.de/opus/volltexte/2021/13452>.
- [Dou21] Gaëtan DOUÉNEAU-TABOT. « Pebble Transducers With Unary Output ». In : *46th International Symposium on Mathematical Foundations of Computer Science, MFCS 2021*. 2021.
- [Kup21] Denis KUPERBERG. « Positive First-order Logic on Words ». In : *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*. 2021, p. 1-13. DOI : [10.1109/LICS52264.2021.9470602](https://doi.org/10.1109/LICS52264.2021.9470602).
- [CVM22] Thomas COLCOMBET, Sam VAN GOOL et Rémi MORVAN. « First-order separation over countable ordinals ». In : *Foundations of Software Science and Computation Structures*. Cham, 2022, p. 264-284.
- [Gou22b] Jean GOUBAULT-LARRECQ. *Non-Hausdorff Topology and Domain Theory. Electronic supplements to the book – Errata*. 2022.
- [BS23] Jason P. BELL et Daniel SMERTNIG. « Computing the linear hull : Deciding Deterministic ? and Unambiguous ? for weighted automata over fields ». In : *2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. Juin 2023, p. 1-13. DOI : [10.1109/lics56636.2023.10175691](https://doi.org/10.1109/lics56636.2023.10175691). URL : <http://dx.doi.org/10.1109/LICS56636.2023.10175691>.
- [Dou23] Gaëtan DOUÉNEAU-TABOT. « Optimization of string transducers ». Thèse de doct. Université Paris Cité, 2023. URL : https://gdoueneau.github.io/pages/DOUENEAU-TABOT_Optimization-transducers_v2.pdf.
- [KNP23] Sandra KIEFER, Lê Thành Dũng NGUYỄN et Cécilia PRADIC. *Refutations of pebble minimization via output languages*. 2023. arXiv : [2301.09234 \[cs.FL\]](https://arxiv.org/abs/2301.09234). URL : <https://arxiv.org/abs/2301.09234>.
- [BK24] Mikołaj BOJAŃCZYK et Bartek KLIN. « Polyregular Functions on Unordered Trees of Bounded Height ». In : *Proceedings of the ACM on Programming Languages* 8. POPL (jan. 2024), p. 1326-1351. DOI : [10.1145/3632887](https://doi.org/10.1145/3632887). URL : <http://dx.doi.org/10.1145/3632887>.
- [DE24] Anuj DAWAR et Ioannis ELEFTHERIADIS. « Preservation Theorems on Sparse Classes Revisited ». In : *49th International Symposium on Mathematical Foundations of Computer Science, MFCS 2024, August 26-30, 2024, Bratislava, Slovakia*. T. 306. 2024, 47 :1-47 :16. DOI : [10.4230/LIPIcs.MFCS.2024.47](https://doi.org/10.4230/LIPIcs.MFCS.2024.47). URL : <https://doi.org/10.4230/LIPIcs.MFCS.2024.47>.
- [Gra24] Thomas GRANDJEAN. *French Cities*. Version 1.0.2. Pour la DREAL Hauts-de-France, service Information, Développement Durable et Évaluation Environnementale, pôle Promotion de la Connaissance. French-cities est un package python de type « boîte à outils » autour des communes et départements français. 2024. URL : <https://pypi.org/project/french-cities/>.

- [PS24] Antoni PUCH et Daniel SMERTNIG. *Factoring through monomial representations : arithmetic characterizations and ambiguity of weighted automata*. 2024. arXiv : [2410.03444v1](https://arxiv.org/abs/2410.03444v1) [math.GR]. URL : <https://arxiv.org/abs/2410.03444v1>.
- [Lop25] Aliaume LOPEZ. « Commutative $\mathbb{N}$ -Rational Series of Polynomial Growth ». In : *42nd International Symposium on Theoretical Aspects of Computer Science (STACS 2025)*. T. 327. Dagstuhl, Germany, 2025, 67 :1-67 :16. DOI : [10.4230/LIPIcs.STACS.2025.67](https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.STACS.2025.67). arXiv : [2404.02232](https://arxiv.org/abs/2404.02232) [cs.LG]. URL : <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.STACS.2025.67>.

A Attestation de réussite



DOCTORAT DE L'UNIVERSITE PARIS-SACLAY

ATTESTATION DE REUSSITE

Le Chef d'établissement de l'Université Paris-Saclay atteste que le diplôme de doctorat de l'université Paris-Saclay, au sein de l'École Doctorale n°580 - Sciences et Technologies de l'Information et de la Communication, dans la spécialité Informatique, a été décerné à

Monsieur Aliaume Dorian Quentin LOPEZ

Né le 27 décembre 1995 à Paris, France

Au titre de l'année universitaire 2022-2023

Titre des travaux : Théorèmes de préservation pour la logique au premier ordre: localité, topologie et constructions limites

Déposés sous le
numéro national de 2023UPASG046
thèse :

Date de soutenance : 12 septembre 2023

Lieu : Salle 127 Bâtiment Olympe de Gouges 8 place Paul Ricoeur 75013 PARIS

Composition du jury

Mme Serenella CERRITO, Professeur des universités, Université d'Evry Val d'Essonne - Président

M. Anuj DAWAR, Professeur des universités, University of Cambridge - Rapporteur

M. Patrice OSSONA DE MENDEZ, Chargé de recherche (HDR), EHESS - Rapporteur

Mme Daniela PETRISAN, Maître de conférences, Université Paris-Cité - Examineur

M. Luca REGGIO, Chargé de recherche, University College London - Examineur

M. Arnaud DURAND, Professeur des universités, Université Paris-Cité - Examineur

Le 20 octobre 2023

Fait à Paris-Saclay

Par délégation du Chef d'établissement de
l'Université Paris-Saclay,
la vice-présidente adjointe chargée du
doctorat, Sylvie Pommier





DOCTORATE OF UNIVERSITE PARIS-SACLAY

CERTIFICATE OF ACHIEVEMENT

The President of Université Paris-Saclay certifies that the doctoral degree from Université Paris-Saclay, within the Doctoral School n°580 - Sciences and Technologies of Information and Communication, in the specialty Informatics, was awarded to

Mr Aliaume Dorian Quentin LOPEZ

Born on December 27th, 1995 in Paris, France

For the academic year 2022-2023

Title of the thesis: First Order Preservation Theorems in Finite Model Theory: Locality, Topology, and Limit Constructions

Legal deposit under
the national thesis
number: 2023UPASG046

Date of defense: September 12th, 2023

Place of defense: Salle 127 Bâtiment Olympe de Gouges 8 place Paul Ricoeur 75013 PARIS

Composition of the jury

Mme Serenella CERRITO, Professeur des universités, Université d'Evry Val d'Essonne - President

M. Anuj DAWAR, Professeur des universités, University of Cambridge - Reviewer

M. Patrice OSSONA DE MENDEZ, Chargé de recherche (HDR), EHESS - Reviewer

Mme Daniela PETRISAN, Maître de conférences, Université Paris-Cité - Examiner

M. Luca REGGIO, Chargé de recherche, University College London - Examiner

M. Arnaud DURAND, Professeur des universités, Université Paris-Cité - Examiner

Done on October 20th, 2023
in Paris-Saclay

As delegated by the President of the
University of Paris-Saclay
the deputy Vice-President for doctoral
research, Sylvie Pommier



B Procès-verbal de soutenance

LOPEZ Aliaume

NNT : 2023UPASG046

Année universitaire 2022-2023

1/6



DOCTORAT DE L'UNIVERSITE PARIS-SACLAY

Documents de soutenance

Le directeur de thèse remet les documents de soutenance, ci-dessous, au président de jury le jour de la soutenance.

Le directeur de thèse **se charge de récupérer les documents signés** et les procurations :

- Il les dépose en **format numérique dans son espace ADUM** à partir du lien prévu à cet effet (le cas échéant, il les transmet par mail au service scolarité)

- **Et** il envoie les originaux par courrier à l'adresse ci-dessous :

Maison du Doctorat-Université Paris-Saclay
Documents de soutenance
ENS Paris-Saclay
4 avenue des sciences
91190 Gif-Sur-Yvette

Nous vous rappelons que l'envoi **des documents originaux** est indispensable après avoir transmis les documents scannés.

LOPEZ Aliaume

NNT : 2023UPASG046

Année universitaire 2022-2023

2/6



DOCTORAT DE L'UNIVERSITÉ PARIS-SACLAY
PROCÈS-VERBAL DE SOUTENANCE DE THÈSE

A compléter et à déposer au service de scolarité à l'issue de la soutenance

Nom et Prénom : LOPEZ Aliaume
Diplôme : Informatique
Titre de la thèse : Théorèmes de préservation pour la logique au premier ordre: localité, topologie et constructions limites.
École Doctorale : Sciences et Technologies de l'Information et de la Communication
Directeur(s) de thèse : Jean GOUBAULT-LARRECQ
Lieu de soutenance : Bâtiment Olympe de Gougues 8 place Paul Ricoeur 75013 PARIS
Date et heure : 12 septembre 2023 à 14h00
Soutenance ☒ PUBLIQUE ☐ À HUIS-CLOS
Cotutelle de thèse : ☐ OUI ☒ NON

Président du jury : Serey Nello (nom, prénom à préciser, date et signature)
CERRIO

12/09/2023
Serey Nello

Le jury prononce:

☒ L'admission au grade de doctorat, sous réserve de finalisation du dépôt légal de la thèse (voir avis du jury sur la finalisation du dépôt légal de la thèse)

☐ L'ajournement

Après avoir prononcé la décision d'admission, le président de jury invite le docteur à prêter serment^{*} :

"En présence de mes pairs.

Parvenu à l'issue de mon doctorat en 'Informatique', et ayant ainsi pratiqué, dans ma quête du savoir, l'exercice d'une recherche scientifique exigeante, en cultivant la rigueur intellectuelle, la réflexivité éthique et dans le respect des principes de l'intégrité scientifique, je m'engage, pour ce qui dépendra de moi, dans la suite de ma carrière professionnelle quel qu'en soit le secteur ou le domaine d'activité, à maintenir une conduite intégrée dans mon rapport au savoir, mes méthodes et mes résultats."

"In the presence of my peers.

Having completed my doctorate in 'Informatics', and having thus practiced, in my quest for knowledge, the exercise of a demanding scientific research, by cultivating intellectual rigor, ethical reflexivity and respect for the principles of scientific integrity, I commit myself, for what will depend on me, in the continuation of my professional career, whatever the sector or the field of activity, to maintain an honest conduct in my relationship to knowledge, my methods and my results."

☒ A l'issue de la soutenance Monsieur LOPEZ Aliaume a prêté serment

^{*} Selon l'article 19bis de l'arrêté du 25 mai 2016 fixant le cadre national de la formation et les modalités conduisant à la délivrance du diplôme national de doctorat, modifié par l'arrêté du 26 août 2022, « A l'issue de la soutenance et en cas d'admission, le docteur prêtre serment, individuellement en s'engageant à respecter les principes et exigences de l'intégrité scientifique dans la suite de sa carrière professionnelle, quel qu'en soit le secteur ou le domaine d'activité ».

LOPEZ Aliaume

NNT : 2023UPASG046

Année universitaire 2022-2023

3/6

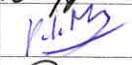
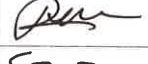
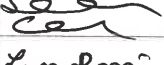
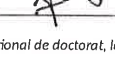
Direction de la thèse et, le cas échéant, co-encadrement du doctorant ou de la doctorante :

Prénom NOM	Rôle	Grade	Affiliations
Jean GOUBAULT-LARRECQ	Directeur de thèse	PREX	UR : Laboratoire Méthodes Formelles Etablissement : ENS Cachan
Sylvain SCHMITZ	Co-encadrant de thèse	MCF	UR : Institut de Recherche en Informatique Fondamentale Etablissement : Université Paris-Cité

La Direction de thèse ne signe en aucun cas le procès-verbal de soutenance. Pour être valides, les documents de soutenance ne doivent pas être modifiés.

Désignation des membres du jury avec voix délibérative

Les membres du jury attestent avoir pris part à la décision d'admission ou d'ajournement. Ils attestent que le directeur ou la directrice de thèse, ainsi que toute autre personne ayant participé à la direction de la thèse ou l'encadrement du doctorant ou de la doctorante, n'a pas pris part à la décision.

Civilité, NOM, Prénom	Fonction	Titre	Visio conférence**	Signature
M. Anuj DAWAR	Rapporteur ☐ Président*	Professeur des universités	NON	
M. Patrice OSSONA DE MENDEZ	Rapporteur ☐ Président*	Chargé de recherche	NON	
Mme Daniela PETRISAN	Examineur ☐ Président*	Maître de conférences	NON	
Mme Serenella CERRITO	Examineur ☒ Président*	Professeur des universités	NON	
M. Luca REGGIO	Examineur ☐ Président*	Chargé de recherche	NON	
M. Arnaud DURAND	Examineur ☐ Président*	Professeur des universités	NON	

* Selon l'article 18 de l'arrêté du 25 mai 2016 fixant le cadre national de la formation et les modalités conduisant à la délivrance du diplôme national de doctorat, les membres du jury désignent parmi eux un président.

Le président du Jury doit être professeur des universités ou assimilé.

« Le directeur de thèse, ainsi que toute autre personne ayant participé à la direction de la thèse, ne prend pas part à la décision. »

** Le président du jury signe le PV pour chacun des membres du jury en visioconférence et joint les procurations signées.

LOPEZ Aliaume

NNT : 2023UPASG046

Année universitaire 2022-2023

4/6



DOCTORAT DE L'UNIVERSITÉ PARIS-SACLAY

RAPPORT DE SOUTENANCE

Nom et prénom du doctorant : LOPEZ Aliaume

Titre de la thèse : Théorèmes de préservation pour la logique au premier ordre: localité, topologie et constructions limites.

École Doctorale : Sciences et Technologies de l'Information et de la Communication

Date de la soutenance : 12 septembre 2023

Président du jury Serenella CERRITO (À COMPLÉTER)

Membres du jury avec voix délibérative :

Nom	Signature	Nom	Signature
Anuj DAWAR		Patrice OSSONA DE MENDEZ	
Daniela PETRISAN		Serenella CERRITO	
Luca REGGIO		Arnaud DURAND	

Les membres du jury attestent avoir pris connaissance de l'intégralité du rapport.

Ils attestent que le directeur ou la directrice de thèse, ainsi que toute autre personne ayant participé à la direction de la thèse ou l'encadrement du doctorant ou de la doctorante, n'a pas pris part à la décision.

Si le rapport comporte plusieurs pages ou s'il est rédigé sur un document distinct, il devra être paraphé sur chaque page et signé par le Président du jury.

LOPEZ Aliaume

NNT : 2023UPASG046

Année universitaire 2022-2023

5/6

université
PARIS-SACLAY

AVIS DU JURY SUR LA FINALISATION
DU DEPOT LEGAL DE LA THESE

La thèse est une œuvre de l'esprit protégée par le droit d'auteur, qui est sécurisée par le dépôt légal de la thèse.

Titre de la thèse : Théorèmes de préservation pour la logique au premier ordre: localité, topologie et constructions limites.

Nom et prénom de l'auteur : LOPEZ Aliaume

Date de soutenance : 12 septembre 2023

Membres du jury avec voix délibérative : M. Anuj DAWAR

M. Patrice OSSONA DE MENDEZ

Mme Daniela PETRISAN

Mme Serenella CERRITO

M. Luca REGGIO

M. Arnaud DURAND

Thèse confidentielle : non

Embargo sur la diffusion : non

(durée de l'embargo limitée à 5 ans, avec prolongation possible, sur demande motivée et justifiée de l'auteur à fournir au plus tôt un an après la soutenance)

À l'issue de la soutenance, le jury estime que la thèse ci-dessus mentionnée :

☒ Le dépôt légal peut être finalisé, après des modifications mineures sur la thèse
Des modifications mineures seront toujours nécessaires, au moins pour mentionner sur la page de couverture le nom du président du jury.

☐ Le dépôt légal ne peut être finalisé qu'après des corrections majeures à effectuer par l'auteur dans un délai de trois mois et après validation par le Président du jury.

Si des corrections majeures sont demandées, identification du membre du Jury désigné par le président pour vérifier les corrections :

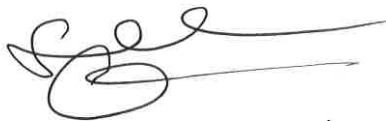
Nom, prénom, titre et fonction dans le jury :

..... (A compléter)

Dans ce cas, le président du jury complète, date et signe le formulaire de vérification des corrections majeures apportées à la thèse et le remet au membre du jury chargé de la vérification.

Nom, prénom, date et signature du Président du jury :

Serenella CERRITO



12/09/2023

LOPEZ Aliaume

NNT : 2023UPASG046

Année universitaire 2022-2023

6/6

**DEMANDE DE VERIFICATION DES CORRECTIONS MAJEURES
A APPORTER A LA THESE***A retourner au service de scolarité daté et signé*

Le président du jury de soutenance de

Nom et Prénom : LOPEZ Aliaume
Diplôme : Informatique
Titre de la thèse : Théorèmes de préservation pour la logique au premier ordre: localité, topologie et constructions limites.
École Doctorale : Sciences et Technologies de l'Information et de la Communication
Lieu de soutenance : Bâtiment Olympe de Gouges 8 place Paul Ricoeur 75013 PARIS
Date et heure : 12 septembre 2023 à 14h00

Désigne :

(À COMPLÉTER) : Nom, Prénom, Titre et fonction dans le Jury

Pour vérifier les corrections majeures de la thèse, demandées par le jury, dans un délai d'un mois à compter de la réception de la thèse révisée.

Date :

Signature du Président du jury (indiquer le nom et prénom)

En tant que membre du jury de soutenance désigné par le président du jury, j'atteste que j'ai vérifié les corrections qui ont été réalisées par le doctorant sur sa thèse, dont la version révisée m'a été remise

Date :

- ☐ Après vérifications, le dépôt légal de la thèse révisée peut être effectué
☐ Après vérifications, le dépôt légal de la thèse révisée ne peut pas être finalisé

Le à

Nom et prénom du membre du jury désigné pour effectuer les vérifications et signature

C Rapport de thèse 1

Anuj Dawar
Professor of Logic and Algorithms



UNIVERSITY OF
CAMBRIDGE

Computer Laboratory

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom

Vice-President for Doctoral Research
Université Paris-Saclay
France.

phone +44 1223 334408
fax +44 1223 334678
email anuj.dawar@cl.cam.ac.uk
date 15 August, 2023

Thesis of Aliaume Lopez

Dear Prof. Pommier,

I am pleased to write in response to your request for an evaluation of the thesis submitted by Mr Aliaume Lopez towards obtaining a doctorate in Informatics.

The thesis *First Order Preservation Theorems in Finite Model Theory: Locality, Topology and Limit Constructions* is a contribution to the subject of finite model theory. This subject seeks to study the expressivity of formal logical languages in the context in which they are interpreted in finite structures. The subject has its origins in the theoretical study of database query languages. I should note that the present thesis contains an excellent motivational chapter at the beginning that links the theoretical questions studied in the thesis to practical considerations that arise naturally from the process of modelling databases as finite relational structures.

Within the context of finite model theory, there is a long history of the study of preservation theorems. Indeed, the study of preservation theorems goes back further as they are a classical element of *model theory*, i.e. the study of the expressive power of logical formalism in the realm of infinite mathematics. The classical preservation theorems, such as the results of Łoś, Tarski and Lyndon, assert that certain semantic limitations on first-order formulas (or theories) correspond naturally to syntactic limitations. As the subject of finite model theory developed its own methods and techniques in the 1980s, it was established that most such preservation theorems do not survive a restriction to finite structures. Indeed, the failure of such preservation results in the finite was one of the early pillars of work in finite model theory. Later, in the early years of the present century, another strand of work sought to investigate the status of these preservation theorems not in the class of *all* finite structures but in restriction to certain well-behaved classes. Many of these so-called *tame* classes are obtained by imposing *sparsity* conditions on the class of structures.

Thus, the study of preservation theorems is an important and recurring theme within the study of finite model theory. There is a large number of both positive and negative results in this domain and an equal variety of methods of establishing them. The negative results rely on building sometimes intricate counter-examples. The positive results have relied on the locality of first-order logic on finite structures. However, it seems this has been up until now an ad hoc collection of methods that work. The present thesis is, to the best of my knowledge, the first attempt at a systematic theory of what makes certain preservation results on well-behaved classes of finite structures possible. In making the step from a collection of disparate, yet significant, results to a more coherent theory, it marks a substantive contribution to our knowledge.

The theory developed leverages the language of topology to express the common elements of various positive preservation theorems. This allows for a single framework in which a number of strands of preservation results, to do with fragments of existential logic, can be expressed. It gives a clear separation between the purely syntactic and the more structural elements of the proofs. The theory is not all encompassing (for instance, Rossman's famous finite homomorphism preservation theorem does not fit so neatly into it) but it is wide in its sweep and convincing in showing that it has some of the right abstractions. The connections it exposes between various results are deep and systematic. It may be argued that the approach has not yet yielded breakthroughs in the form of major new preservation theorems, but I find that the systematic treatment is truly insightful. Moreover, the algebraic treatment of classes of structures with closure properties for preservation theorems does yield infinitely new well-behaved classes.

The thesis contains seven substantial chapters (not counting the opening chapter of preliminaries). The first two form an introduction to the subject, with Chapter 2 providing the motivation from considerations of databases, and Chapter 3 giving a background in the form of a survey of known results about preservation theorems in the realm of both finite and general model theory. Already this survey chapter contains substantial new insights as it gives a unified treatment of these results and establishes the common topological framework within which they are treated. This topological framework then forms the basis of the work in the chapters that follow. Chapter 4 explores the connection between locality and preservation, establishing a new preservation theorem for existential local sentences which, unsurprisingly, fails in the finite. Chapter 5 pushes the analysis of locality and preservation further, showing how preservation for a class of structures can be established by considering just the class of local neighbourhoods occurring in the structures. The topological considerations that arise from this study are then extended in the subsequent chapters. Chapter 6 considers algebraic constructions of classes of structures from which new classes exhibiting preservation properties can be derived. This leads in particular to limit constructions from infinite sequences and orders. The subsequent Chapters 7 and 8 develop the topological machinery on words and trees much further. This development of new topological results is motivated by the work on preservation but by the end is somewhat far removed from it. Overall, there is an impressive amount of new mathematical material in the thesis.

Let me now make a few general remarks about the thesis. The first is that the combination of topological methodology and the subject matter of finite model theory is highly original. This is a marker of the novelty of the work and the creativity of the candidate. The second remark is that the exposition is of exceptionally high quality. The thesis is a pleasure to read. It not only shows a mastery of the technical material (which itself is necessary in order to give such an original exposition), but a high degree of skill in scientific communication which is itself a highly desirable quality in a researcher. It is clear that a great deal of thought has gone into the exposition. A third remark is that much of the research presented in the thesis has been published in high quality publications in excellent, selective venues. The two papers at the CSL 2021 and LICS 2022 conferences were single authored by the candidate which demonstrates that the work is the result of independent research.

In summary, this thesis is an excellent piece of work and I clearly recommend that the candidate be allowed to proceed to a public defence. I have a list of minor corrections that I will make available directly to the candidate.

Yours sincerely,



Prof. Anuj Dawar

D Rapport de thèse 2



Centre d'Analyse et de Mathématiques Sociales

École des Hautes Études en Sciences Sociales
Centre National de la Recherche Scientifique UMR 8557
54 boulevard Raspail — 75006 Paris — FRANCE

14 août 2023

Patrice Ossona de Mendez

Chargé de Recherche CNRS (HDR)
Invited Professor at Charles University, Prague
Associated researcher at Zhejiang Normal University
Editor-in-Chief of the *European Journal of Combinatorics*

Rapport concernant la thèse intitulée

« **First Order Preservation Theorems in Finite Model Theory : Locality, Topology, and Limit Constructions** »

présentée par M. Aliaume LOPEZ

en vue d'obtenir le grade de docteur ès sciences (spécialité : informatique) de l'Université Paris-Saclay.

Les théorèmes de préservation de la logique du premier ordre forment un corpus classique de résultats en Théorie des Modèles. Il s'agit notamment des théorèmes suivants :

- *Théorème de Łoś–Tarski* : Un énoncé est préservé par extensions si, et seulement si, il est logiquement équivalent à un énoncé existentiel.
- *Théorème de positivité de Lyndon* : Un énoncé est préservé par homomorphismes surjectifs si, et seulement si, il est logiquement équivalent à un énoncé positif.
- *Théorème de préservation par homomorphismes* : Un énoncé est préservé par homomorphismes si, et seulement si, il est logiquement équivalent à un énoncé existentiel positif.

Ces théorèmes de préservation établissent une équivalence entre une propriété *sémantique* (préservation de la satisfaction lors de certaines opérations) et une propriété *syntactique* (existence d'un énoncé équivalent ayant une certaine forme).

Ces énoncés classiques sont valides lorsque la propriété de préservation est vérifiée par toutes les structures (de signature donnée), qu'elles soient finies ou infinies. Il est donc naturel de considérer des hypothèses plus faibles (préservation d'une propriété sémantique par les structures finies ou par les structures appartenant à une classe $\mathcal{C}$ de structures finies) en contrepartie d'une conclusion plus faible (équivalence matérielle avec un énoncé sur l'ensemble des structures finies ou sur une classe $\mathcal{C}$ de structures finies). On parle alors de *relativisation* des théorèmes de préservation. L'étude de ces relativisations (de leur validité ou non selon la classe $\mathcal{C}$ de structures finies) est particulièrement difficile et a été essentiellement abordée au cas par cas.

L'objet principal de la thèse de M. Aliaume LOPEZ est de proposer un cadre conceptuel et théorique permettant de déterminer la validité (ou non) de la relativisation de théorèmes de préservation à des classes de structures finies.

Dans un premier temps, Aliaume LOPEZ considère comment les propriétés de localité de la logique du premier ordre peuvent être utilisées pour démontrer des théorèmes de préservation. À cette occasion, il étudie les énoncés du premier ordre préservés par extension élémentaires locales, et démontre un nouveau théorème de préservation :

Sur une classe $\mathcal{C}$ de structures relationnelles finies, un énoncé φ est équivalent, sur $\mathcal{C}$, à un énoncé existentiel local si, et seulement si, il est équivalent, sur $\mathcal{C}$, à une combinaison booléenne positive d'énoncés basiques-locaux.

Ce résultat a fait l'objet d'une présentation au *Symposium on Logic in Computer Science* (LICS) en 2022.

Aliaume LOPEZ identifie le rôle joué par la propriété de bel ordre dans les démonstrations de relativisation de théorèmes de préservation. Afin de généraliser les résultats connus et d'offrir un cadre conceptuel pour obtenir de nouvelles relativisations, Aliaume LOPEZ considère une version topologique de la structure bel ordre, l'espace noethérien, qui est un espace topologique sans suite strictement croissante infinie d'ouverts.

Aliaume LOPEZ considère alors une représentation topologique des énoncés de préservation basée sur des *espaces préspectraux logiquement présentés*, qui sont des espaces topologiques sur des classes de structures finies, dont les ouverts compacts forment une algèbre de Boole qui engendre la topologie. Aliaume LOPEZ montre ensuite comment des preuves de relativisation de théorèmes de préservation s'appuie sur une définition inductive des topologies considérées et, plus spécifiquement, sur une définition par plus petit point fixe d'une fonction de raffinement.

Conséquemment, Aliaume LOPEZ introduit la notion d'expandeur de topologie, qui est une fonction E monotone sur l'espace des topologies, qui préserve le caractère noethérien et qui est compatible avec l'opération de trace sur un fermé (dans le sens que, pour toute topologie τ raffinée par E , la trace de l'image de τ sur un fermé H de τ coïncide avec la trace sur H de l'image de la trace sur H de τ). Aliaume LOPEZ démontre alors le résultat principal de cette seconde partie :

Si τ est une topologie noethérienne raffinée par un expandeur de topologie E , alors le plus petit point fixe de E au dessus de τ est une topologie noethérienne.

À noter que la preuve de ce théorème s'appuie sur un argument de type "mauvaise séquence minimale" inspiré de ceux qui apparaissent dans la démonstration du lemme d'Higman, du théorème de Kruskal, ou de leurs extensions topologiques par Goubault-Larrecq.

Aliaume LOPEZ considère alors un point de vue catégorique basée sur la notion de foncteur analytique de Hasegawa et sur l'utilisation d'une algèbre initiale associée. Cette approche permet de retrouver les notions de support d'un objet et d'un ordre des sous-structures implicitement présent dans le travail de Hasegawa. Aliaume LOPEZ définit ensuite un expandeur de topologie basé sur les sections finissantes de cet ordre et démontre que si le relèvement d'un foncteur analytique G qui préserve les plongements topologiques définit sur toute algèbre initiale de G une topologie noethérienne (qu'il dénomme *topologie de divisibilité*), qui est le plus petit point fixe de l'expandeur de topologie associé à l'ordre des sous-structures.

Appréciation.

Les résultats présentés par Aliaume LOPEZ dans sa thèse sont particulièrement importants et difficiles. Ils témoignent d'une forte intuition de l'approche topologique et catégorique de la théorie des modèles finis et des propriétés essentielles à la définition de certaines constructions. Il ne fait aucun doute que cette compréhension intime sera un atout majeur important pour Aliaume LOPEZ dans ses recherches futures, et qu'elle lui permettra d'être un moteur dans l'équipe qui aura la chance de l'accueillir.

En conséquence, je suis très favorable à la soutenance de la thèse de M. Aliaume LOPEZ, qui est sans aucun doute promis à une brillante carrière de chercheur.

Cordialement,

Patrice Ossona de Mendez



E Rapport de soutenance

Rapport de soutenance de la thèse de doctorat de M. Lopez

Dans le cadre de ses délibérations, le jury a apprécié la qualité des travaux du doctorant Alliaume LOPEZ, leur caractère novateur, l'aptitude du doctorant à les situer dans leur contexte scientifique ainsi que ses qualités d'exposition.

Le candidat a présenté ses contributions en informatique fondamentale autour des théorèmes de préservation en logique. Il s'agit de résultats originaux, profonds et conceptuellement difficiles sur un sujet à l'interface de la théorie des modèles finis et de la topologie. Les contributions du candidat portent non seulement sur chacun de ces deux domaines mais établissent aussi des liens nouveaux et surprenants entre ceux-ci.

L'exposé a montré une excellente maîtrise du sujet de la part de M. Lopez et ceci lui a permis de présenter des exemples concrets illustrant des concepts techniques abstraits et complexes et de donner des motivations aux résultats théoriques. La présentation orale, bien pédagogique, a d'ailleurs renforcé l'excellente impression déjà laissée aux membres du jury par le manuscrit de mémoire, qui est remarquablement bien écrit.

Le jury a posé plusieurs questions et les réponses données ont montré non seulement une compréhension profonde du sujet d'étude, mais aussi l'ouverture intellectuelle du candidat, qui avait évidemment déjà réfléchi à des questions allant au-delà des résultats présentés dans le mémoire.

Les conclusions du mémoire et de la soutenance orale ont présenté des perspectives de recherche ambitieuses et stimulantes.

Pour l'ensemble de ces raisons et après en avoir délibéré, le jury prononce à l'unanimité l'admission de M. Lopez au grade de docteur de l'université Paris-Saclay en Informatique et l'invite à prononcer le serment des docteurs.

Le 12/09/2023

La présidente du jury

Serenella Cerrito



SC

F Ackermann Award 2024

The Ackermann Award 2024

Maribel Fernández 

Department of Informatics, King's College London, UK

Prakash Panangaden 

McGill University, Canada

Abstract

Report on the 2024 Ackermann Award.

2012 ACM Subject Classification Theory of computation → Logic and verification; Theory of computation → Finite Model Theory; Theory of computation → Proof theory; Theory of computation → Transducers

Keywords and phrases finite automaton, string transducer, class membership problem, first-order logic, preservation theorem, finite model theory

Digital Object Identifier 10.4230/LIPICs.CSL.2025.1

Category Ackermann Award

Introduction

The Ackermann Award is the **EACSL Outstanding Dissertation Award for Logic in Computer Science**. It is presented at CSL, the annual conference of the EACSL (European Association for Computer Science Logic). This year the 20th Ackermann Award is presented at CSL 2025 in Amsterdam, The Netherlands.

A call for nominations was issued in February 2024, open to any PhD dissertation (on any topic represented at the annual CSL and LICS conferences) formally accepted by a degree-granting institution in fulfilment of the PhD degree between 1 January 2023 and 31 December 2023.

The Jury received ten nominations, which came from a number of different countries around the world: the nominees obtained their doctorates at institutions in Belgium, Canada, Czech Republic, France, Poland and the United Kingdom. The topics covered a wide range of areas in Logic and Computer Science.

This year we received a particularly strong set of nominations. All the nominated PhD theses contained significant contributions to their particular fields. On behalf of the Ackermann Jury, we extend our warmest congratulations to all the nominated candidates for their outstanding work.

All the submissions were evaluated by the Jury, and after two phases of reviewing and extensive discussion, the jury decided to grant the **2024 Ackermann Award** jointly to (in alphabetic order):

- **Gaëtan Douéneau-Tabot** for the PhD thesis entitled *Optimization of string transducers*, completed at *University Paris-Cité*, France, in 2023;
- **Aliaume Lopez** for the PhD thesis entitled *First Order Preservation Theorems in Finite Model Theory: Locality, Topology, and Limit Construction*, completed at *University Paris-Saclay*, France, in 2023.



© Maribel Fernández and Prakash Panangaden;

licensed under Creative Commons License CC-BY 4.0

33rd EACSL Annual Conference on Computer Science Logic (CSL 2025).

Editors: Jörg Endrullis and Sylvain Schmitz; Article No. 1; pp. 1:1–1:5



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1:2 The Ackermann Award 2024

Citation for Gaëtan Douéneau-Tabot

Gaëtan Douéneau-Tabot shares the *2024 Ackermann Award* of the European Association of Computer Science Logic for the PhD thesis

Optimization of string transducers,

which solves open class membership problems for various transducer models, providing effective membership procedures that give rise to program optimization techniques. To achieve this goal, the thesis develops an extensive toolbox for solving class membership problems for transducers, including new characterisations of transducer behaviours.

Background to the thesis

This thesis deals with transducer models, that is, finite automata extended with string outputs, which define functions from strings to strings. String transducers are finite-state devices that represent programs with bounded memory. They are used in language processing tools, compilers, streaming algorithms and model-checkers, amongst others.

Various transducer models have been defined as extensions of the basic model by adding features to increase their expressive power. A natural question then arises: what is the expressive power of each class of transducers? This can be rephrased as a class membership problem: given a transducer with complex features, is there a simpler transducer that has the same behaviour? Such a question can be interpreted as a program optimisation problem: given a program, can we build an equivalent program that requires less resources?

Class membership problems in this context are known to be challenging. In his PhD thesis, Douéneau-Tabot solves multiple instances of the problem, obtaining new results that characterise various classes of transducer models and help understand their relationships.

Contributions of the thesis

The thesis solves open class membership problems for various transducer models. In each case the membership procedure is effective, in the sense that it builds a more efficient transducer whenever one exists, thus providing program optimisation techniques in the setting of transducers. In particular, the techniques can be applied to automatically remove recursion or nested loops for specific classes of programs.

In addition, the thesis provides new computation models and new characterisations that capture pre-existing classes of transductions. These results provide new insights on the expressive power of different kinds of transducers.

The proof techniques introduced are also valuable as a generic toolbox for solving other class membership problems in transducer models, and more generally in automata theory.

Biographical sketch

Gaëtan Douéneau-Tabot carried out his PhD studies at University Paris-Cité, under the supervision of Olivier Carton and Emmanuel Filiot, from 2020 to 2023. During his PhD he (co)-authored papers published in the proceedings of conferences such as ICALP 2023, LICS 2023, FoSSaCS 2023, MFCS 2022, ICALP 2022, MFCS 2021, MFCS 2020. His MFCS 2022 and ICALP 2022 papers won “best student paper” awards.

M. Fernández and P. Panangaden

1:3

Citation for Aliaume Lopez

Aliaume Lopez shares the *2024 Ackermann Award* of the European Association of Computer Science Logic for the PhD thesis

First Order Preservation Theorems in Finite Model Theory: Locality, Topology, and Limit Constructions,

which presents a systematic approach to investigating preservation theorems in Finite Model Theory, providing tools to explain when and why some classes of structures are well-behaved with respect to preservation theorems. The approach provides a compositional theory for preservation theorems that was previously lacking.

Background to the thesis

Preservation theorems in first-order logic are a collection of results derived from classical model theory, which establish a direct correspondence between the semantic properties of formulas and the constraints imposed on their syntax.

These theorems have a practical impact in computer science, where they can be used for example to characterise syntactic classes of database queries for which the termination and correctness of database algorithms are guaranteed. Unfortunately preservation theorems are notably challenging when focusing on finite models – the models considered in computer science applications. Identifying well-behaved classes has been an active domain of research for the last sixty years, with a series of negative results as well as some positive ones, which motivated the quest for a systematic approach to the problem.

Contributions of the thesis

This thesis presents a systematic approach to investigating preservation theorems within the realm of Finite Model Theory. The traditional ad-hoc proofs are replaced with a theoretical framework that generalises techniques based on locality, and introduces a topological presentation of preservation theorems called logically presented pre-spectral spaces.

Introducing these topological spaces enables the development of a compositional theory for preservation theorems. Additionally, this thesis develops a methodology to systematically examine preservation theorems across inductively defined classes of finite structures, by proving a generic fixed point theorem for a topological restriction of logically presented pre-spectral spaces: Noetherian spaces, which are topological spaces in which every open set is compact.

Biographical sketch

Aliaume Lopez carried out his PhD studies under the supervision of Jean Goubault-Larrecq (École Normale Supérieure Paris-Saclay) and Sylvain Schmitz (University Paris-Cité). The thesis is built around three articles published at CSL 2021, LICS 2022 and FoSSaCS 2023 (he published other papers during his PhD unrelated with the thesis). He is the winner of the E.W. Beth Dissertation Prize 2024. Currently he is a postdoctoral researcher at the University of Warsaw.

1:4 The Ackermann Award 2024**Jury**

The jury for the **Ackermann Award 2024** consisted of nine members, two of them *ex officio*, namely, the president and the vice-president of EACSL. In addition, the jury also included a representative of SIGLOG (the ACM Special Interest Group on Logic and Computation).

The members of the jury were:

- Albert Atserias (Technical University of Catalonia);
- Christel Baier (Technical University Dresden);
- Andrej Bauer (University of Ljubljana);
- Maribel Fernández (King's College London), president of EACSL;
- Joost-Pieter Katoen (RWTH Aachen University), ACM SIGLOG representative;
- Delia Kesner (IRIF, University Paris Cité);
- Slawomir Lasota (University of Warsaw);
- Florin Manea (University of Göttingen), vice-president of EACSL;
- Prakash Panangaden (McGill University);

Previous winners

Previous winners of the Ackermann Award were

2005, Oxford:

Mikołaj Bojańczyk from Poland,
Konstantin Korovin from Russia, and
Nathan Segerlind from the USA.

2006, Szeged:

Balder ten Cate from the Netherlands, and
Stefan Milius from Germany.

2007, Lausanne:

Dietmar Berwanger from Germany and Romania,
Stéphane Lengrand from France, and
Ting Zhang from the People's Republic of China.

2008, Bertinoro:

Krishnendu Chatterjee from India.

2009, Coimbra:

Jakob Nordström from Sweden.

2010, Brno:

no award given.

2011, Bergen:

Benjamin Rossman from USA.

2012, Fontainebleau:

Andrew Polonsky from Ukraine, and
Szymon Toruńczyk from Poland.

2013, Turin:

Matteo Mio from Italy.

2014, Vienna:

Michael Elberfeld from Germany.

2015, Berlin:

Hugo Férée from France, and
Mickael Randour from Belgium.

M. Fernández and P. Panangaden**1:5****2016, Marseille:**

Nicolai Kraus from Germany.

2017, Stockholm:

Amaury Pouly from France.

2018, Birmingham:

Amina Doumane from France.

2019, Barcelona (conference in 2020):

Antoine Mottet from France.

2020, Ljubljana (conference online in 2021)

Benjamin Kaminski from Germany.

2021, Göttingen (conference online in 2022)

Marie Fortin from France, and

Sandra Kiefer from Germany.

2022, Warsaw (conference in 2023)

Alexander Bentkamp from The Netherlands.

2023, Naples (conference in 2024)

Gabriele Vanoni from Italy.

Detailed reports on their work appeared in the CSL proceedings and are also available on the EACSL homepage.