2013–2016	→	2016–2018	→	2018–2022	→	2023	→	2024–2026 ·······
Baccalauréat Montréal		Maîtrise Montréal		Doctorat Coimbra		Postdoc Paris		Postdoc Prague

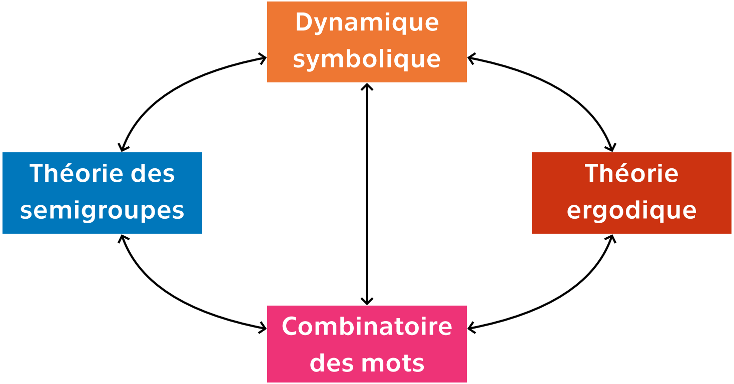
Aspects combinatoires, algébriques et ergodiques de la dynamique symbolique

Université de Moncton

Herman Goulet-Ouellet

4 avril 2025

Vue d'ensemble



Partie 1

Dynamique symbolique

Dynamique symbolique

Espaces symboliques: espaces de mots infinis vus comme systèmes dynamiques.

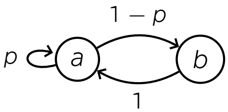
$$
\begin{aligned}
&abaababaabaababaababaabaa\ldots\\
&baababaabaababaababaabaab\ldots\\
&aababaabaababaababaabaaba\ldots
\end{aligned}
$$

- Origine dans l'approximation discrète de systèmes dynamiques continus (Morse et Hedlund, 1938, 1940).
- Liens avec la combinatoire, l'algèbre, la théorie des probabilités, la géométrie, la logique, la théorie des nombres, etc.

Chaînes de Markov

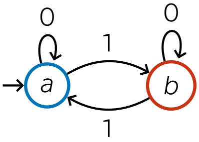
Les **chaînes de Markov** définissent des mesures de probabilité sur les espaces symboliques.

Par exemple, la chaîne de Markov suivante définit une mesure de probabilité sur l'espace symbolique X formé des mots infinis sur $\{a, b\}$ qui évitent le facteur bb.



Mots automatiques

Les **mots automatiques** sont des mots infinis obtenus en lisant les représentations d'entiers (e.g. binaires) dans des automates.

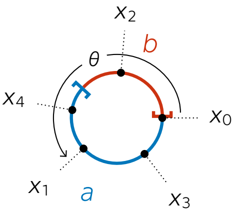


0	1	10	11	100	101	110	111	
a	b	b	a	b	a	a	b	$\ldots$

MOT DE THUE–MORSE. Parité du nombre de 1 dans les représentations binaires des entiers naturels.

Mots sturmiens

Let **mots sturmiens** sont des mots infinis obtenus par certains codages discrets de rotations apériodiques du cercle.



$a \quad b \quad a \quad a \quad b \quad a \quad b \quad a \quad \dots$

MOT DE FIBONACCI. Codage de l'orbite de 0 sous la rotation de longueur $\frac{-1+\sqrt{5}}{2}$ par les intervalles $[0, \frac{3-\sqrt{5}}{2})$, $[\frac{3-\sqrt{5}}{2}, 0)$.

Combinatoire des mots

Espace symbolique
mots infinis

$\longrightarrow$

Langage
mots finis

$$\ldots\ a\ b\ a\ a\ b\ \boxed{a\ b\ a\ a\ b\ a}\ a\ b\ a\ b\ \ldots$$

Problème fondamental

Comprendre les espaces symboliques grâce aux propriétés **combinatoires** et **algébriques** de leurs facteurs finis.

1. Analyser la **complexité en facteurs**.
2. Étudier le comportement des facteurs à l'intérieur de différentes **structures algébriques** (monoïdes ou groupes, par exemple).
3. Se prête bien à une approche expérimentale (Python, SageMath, Gap, etc.).

Mots de retour

Mots de retour: séparations entre les occurrences d'un facteur donné dans les éléments d'un espace symbolique.

$$\dots \underbrace{\overbrace{a\,b}^{u}\,b}_{r_1}\,\underbrace{\overbrace{a\,b}^{u}\,a}_{r_2}\,\underbrace{\overbrace{a\,b}^{u}\,b\,a}_{r_3}\,\underbrace{\overbrace{a\,b}^{u}}_{r_4}\,\underbrace{\overbrace{a\,b}^{u}\,b}_{r_5}\,\underbrace{\overbrace{a\,b}^{u}\,a}_{r_6}\,\underbrace{\overbrace{a\,b}^{u}}_{r_7}\,\overbrace{a\,b}^{u}\dots$$

$$\mathcal{R}_X(ab) = \{abb, aba, abbb, ab\}$$

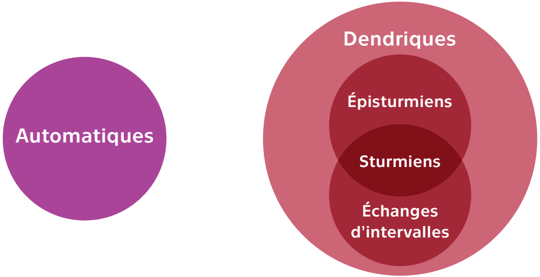
- Notion centrale en dynamique symbolique, utilisée comme outil de classification.
- Liée à plusieurs questions combinatoires (complexité en facteurs, palindromes).

Théorème (Vuillon, 2001)

Un espace symbolique est sturmien $\iff$ $\#\mathcal{R}_x(w) = 2$ pour tout facteur w.

Problèmes de classification

Problème fondamental. Comprendre les espaces symboliques grâce aux propriétés combinatoires et algébriques de leurs facteurs finis.



🛈 Les **espaces dendriques** sont une généralisation combinatoire importante des espaces sturmiens (Berthé et al., 2015). Ils sont disjoints des espaces automatiques.

Caractérisation des dendriques

F. Gheeraert, H. Goulet-Ouellet, J. Leroy, et P. Stas (2025). **Algebraic characterization of dendricity.** *Electron. J. Comb.* 32.1

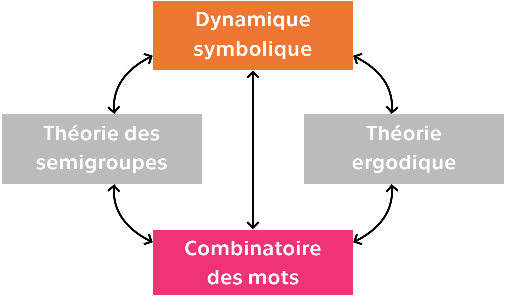
Espace symbolique X $\longrightarrow$ **Mots de retour** $\mathcal{R}_X(w)$ $\longrightarrow$ **Sous-groupes** $\langle \mathcal{R}_X(w) \rangle$

Théorème

Un espace symbolique X est dendrique $\iff$ tous les $\mathcal{R}_X(w)$ forment des bases du groupe libre.

- $\Longrightarrow$: "Théorème du retour" de Berthé et al., 2015.
- $\Longleftarrow$: Gheeraert, Goulet-Ouellet, Leroy, et Stas, 2025.

Récapitulatif



- Espaces symboliques, liens avec les chaînes de Markov.
- Familles d'espaces (sturmiens, automatiques, etc.).
- Combinatoire des mots, mots de retour (approche expérimentale).
- Caractérisation des espaces dendriques.

Partie 2

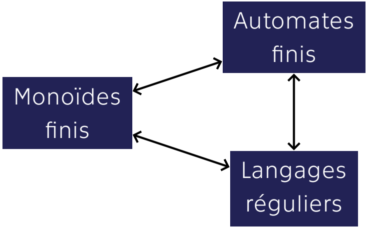
Aspects algébriques

Semigroupes et monoïdes

Définition

- Un **semigroupe** est un ensemble muni d'une opération associative.
- Un **monoïde** est un semigroupe avec un élément neutre.

Par exemple, l'ensemble des mots finis sur un alphabet A, dénoté A^*, forme un monoïde pour la concaténation (monoïde libre).

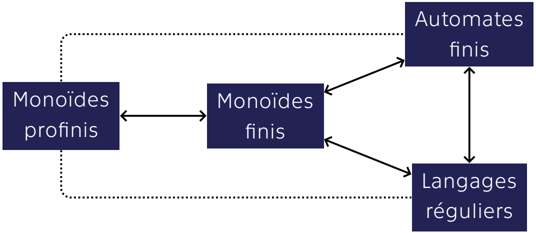


- Les **langages réguliers** (regex) sont décrits par des machines finies appelées **automates**.
- Il y a une correspondance naturelle entre les monoïdes finis et les automates.
- Les monoïdes finis sont intimement liés aux automates et aux langages réguliers.

Monoïdes finis et profinis

Monoïdes profinis: structures **algébriques-topologiques** qui permettent d'encoder le comportement des familles de monoïdes finis.

- Apparaissent dans des contextes très variés, comme les algèbres de Boole, la théorie des modèles, la théorie de Galois, et l'analyse p-adique.
- Permettent d'utiliser des outils topologiques pour étudier les langages réguliers et la combinatoire des mots.



Groupes de Schützenberger

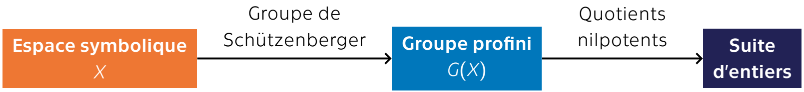
Groupes de Schützenberger: groupes profinis naturellement associés à certains espaces symboliques.

$$\begin{array}{c}\textbf{Espace symbolique}\\ X\end{array} \xrightarrow{\text{Groupe de Schützenberger}} \begin{array}{c}\textbf{Groupe profini}\\ G(X)\end{array}$$

- Étude dynamique des monoïdes profinis (Almeida, 2005).
- La construction de $G(X)$ est liée aux facteurs finis de X.
- $G(X)$ est invariant à isomorphisme près (Costa, 2006).
- Permet de distinguer algébriquement certaines familles (sturmiens, automatiques, etc.)

Invariants nilpotents

H. Goulet-Ouellet (2022). **Pronilpotent quotients associated with primitive substitutions.** *J. Algebra* 606.



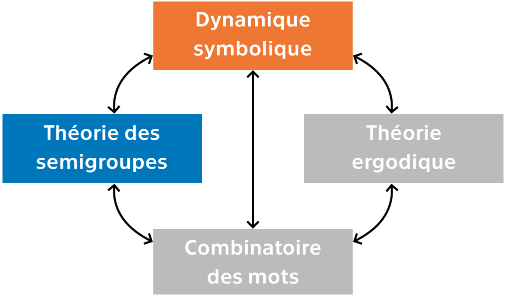
Extraction d'invariants simplifiés (**empreinte nilpotente**) par des méthodes algébriques utilisant les mots de retour $\mathcal{R}_X(w)$.

Théorème

L'empreinte nilpotente suffit à distinguer les espaces dendriques et automatiques.

- L'empreinte nilpotente est encodé par une suite d'entiers simple à calculer.
- Pour Thue–Morse: $1, 2, 2, \dots$.
- Pour Fibonacci: $2, 2, 2, \dots$.

Récapitulatif



- Monoïdes finis et profinis.
- Liens avec les langages réguliers (regex) et automates.
- Construction d'invariants profinis (groupe de Schützenberger).
- Extraction d'invariants simplifiés (empreinte nilpotente).

Partie 3

Aspects ergodiques

Densité des langages

Définition

Soit μ une mesure de probabilité sur un espace symbolique. La **densité** d'un langage $L \subseteq A^*$ relativement à μ est la limite

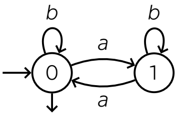
$$\delta_\mu(L) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\{x \in X \mid x_0 \dots x_k \in L\}).$$

Interprétation probabiliste: pour les langages réguliers, la densité correspond à la fréquence moyenne de visites par les marches infinies sur un monoïde fini.

"mots avec un nombre pair de a"

$$L = \{w \in A^* \mid \#_a w \equiv 0 \mod 2\}$$

$$b^*(ab^*ab^*)^*$$



$\mathbb{Z}/2\mathbb{Z}$

Un peu d'histoire

$$\delta_\mu(L) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\{x \in X \mid x_0 \dots x_k \in L\}).$$

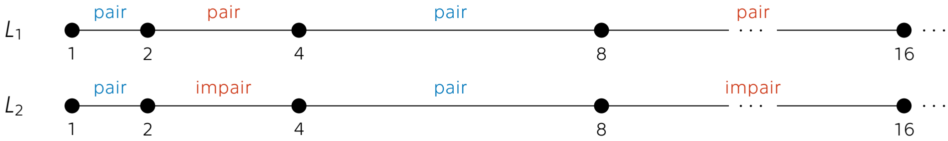
La densité des langages peut être vue comme une généralisation de la densité naturelle en théorie des nombres. Elle a été étudié dans plusieurs contextes.

- 1951 Travaux de **Kakutani** sur les chaînes de Markov (liés indirectement).
- 1969 Travaux de **Veech** sur une variante du théorème de Kronecker–Weyl.
- 1972 **Berstel** étudie les densités pour les mesures de Bernoulli.
- 1989 **Hansel et Perrin** étudient les densités pour les chaînes de Markov.
- 1993 **Lynch** étudie les densités en lien avec la logique du premier ordre.
- 2015 Travaux de **Sin'ya** sur une loi zéro-un pour les langages réguliers.

Existence

$$\delta_\mu(L) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\{x \in X \mid x_0 \dots x_k \in L\}).$$

$$L_1 = \{w \in A^* \mid |w| \equiv 0 \mod 2\},$$
$$L_2 = \{w \in A^* \mid |w| \equiv \lfloor \log_2(|w|) \rfloor \mod 2\}.$$



Dans cet exemple, $\delta_\mu(L_1) = \delta_\mu(L_2) = 1/2$, mais la densité $\delta_\mu(L_1 \cap L_2)$ **n'existe pas**.

Mesures ergodiques

Définition

Soit S la fonction de décalage, $Sx_n = x_{n+1}$. Une mesure de probabilité μ sur un espace symbolique est **ergodique** si

$$\forall B, C, \quad \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(B \cap S^{-n}C) = \mu(B)\mu(C).$$

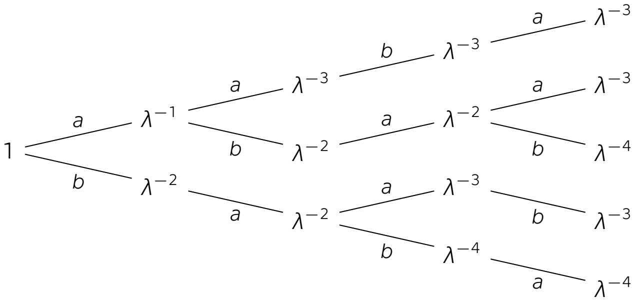
(Tous les évènements sont deux-à-deux asymptotiquement indépendants.)

La **théorie ergodique** est une branche importante de la théorie des systèmes dynamiques, initialement motivée par la physique statistique.

1. Toute les mesures de **Bernoulli** sont ergodiques.
2. Les mesures de **Markov** irréductibles sont ergodiques.

Mesure de Fibonacci

ℹ Les espaces symboliques engendrés par les mots de **Fibonacci** et de **Thue–Morse**, par exemple, admettent une unique mesure de probabilité ergodique (Michel, 1974).



MESURE DE FIBONACCI. Unique mesure ergodique supportée par l'espace symbolique engendré par le mot de Fibonacci *abaababa* $\cdots$ ($\lambda = \frac{1+\sqrt{5}}{2}$).

Densité des langages à groupe

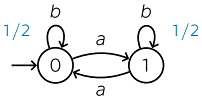
V. Berthé, H. Goulet-Ouellet, C.-F. Nyberg-Brodda, D. Perrin, et K. Petersen (2024). ***Density of group languages in shift spaces.*** Preprint.

$$\delta_\mu(L) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(\{x \in X \mid x_0 \dots x_k \in L\}).$$

Théorème

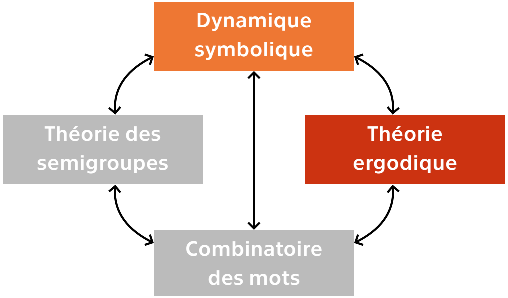
Pour tout **langage à groupe**[1] L et toute **mesure ergodique** μ, la densité $\delta_\mu(L)$ existe.

- **Formules closes** dans plusieurs cas importants.
- **Équidistribution** pour les mesures supportées par les espaces dendriques.



[1] Sous-famille des langages réguliers

Récapitulatif



- Notion de densité.
- Interprétation probabiliste (marches infinis sur des monoïdes).
- Problème de l'existence.
- Ergodicité.
- Théorème d'existence pour les langages à groupes.

Partie 4

Conclusion

Projets de recherche

1. Volet combinatoire

Approfondir l'utilisation des mots de retour comme outil de classfication.

2. Volet profini

Poursuivre l'étude des groupes de Schützenberger.

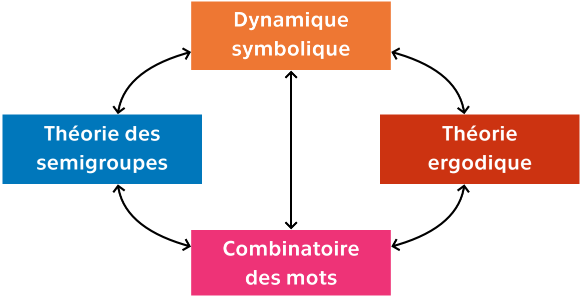
3. Volet ergodique

Étudier les densités au-delà des langages à groupe.

2013–2016 → 2016–2018 → 2018–2022 → 2023 → 2024–2026 ……

2013–2016	2016–2018	2018–2022	2023	2024–2026
Baccalauréat Montréal	Maîtrise Montréal	Doctorat Coimbra	Postdoc Paris	Postdoc Prague

- **7 articles** publiés.
- **4 articles** en cours de révision.
- **5+ articles** en préparation.
- **1 chapitre** de livre en préparation.
- **20+ exposés** (conf, séminaires).



Jorge Almeida Valérie Berthé Alfredo Costa France Gheeraert Karel Klouda
Julien Leroy Dominique Perrin Karl Petersen Štěpán Starosta Pierre Stas

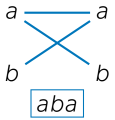
Extra

Pistes de recherche

Volet combinatoire

Bispécial: facteur prolongeable par plusieurs lettres dans un espace symbolique.

$$\begin{array}{l} \dots\ b\ \boxed{a\ b\ a}\ a\ \dots \\ \dots\ a\ \boxed{a\ b\ a}\ a\ \dots \\ \dots\ a\ \boxed{a\ b\ a}\ b\ \dots \end{array}$$



Objectifs

1. Concevoir et implémenter un algorithme pour calculer les facteurs bispéciaux.
    - Premiers résultats soumis récemment (avec K. Klouda et Š. Starosta).
    - Implémentation en SageMath et/ou Python.
2. Développer une approche systématique au calcul des mots de retour.
    - Applications aux deux autres volets.
    - Liens vers d'autres problèmes combinatoire (complexité en facteurs, répétitions).

Volet profini

Substitution. Morphisme $\sigma\colon A^* \to A^*$. L'itération de substitutions permet de construire des mots infinis et des espaces symboliques (Thue–Morse, Fibonacci, etc.).

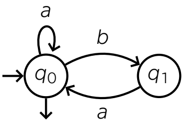
Substitutif	$A^* \xleftarrow{\sigma} A^* \xleftarrow{\sigma} A^* \xleftarrow{\sigma} A^* \leftarrow \cdots$	itération d'une seule substitution
S-adique	$A_0^* \xleftarrow{\sigma_0} A_1^* \xleftarrow{\sigma_1} A_2^* \xleftarrow{\sigma_2} A_3^* \leftarrow \cdots$	itération de plusieurs substitutions

Objectifs

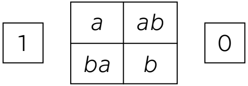
1. Établir une théorie S-adique profinie.
    - Les groupes $G(X)$ demeurent mal compris en dehors du cadre substitutif.
    - Le cadre S-adique offre un degré de généralité beaucoup plus large.
2. Étudier les quotients résolubles des groupes de Schützenberger.
    - Étendre l'idée d'empreinte nilpotente pour obtenir des invariants plus forts.

Volet ergodique

Monoïde de transition: défini par l'action des lettres sur les états d'un automate.



Automate



Monoïde de transition

Objectifs

1. Calculer les densités pour les langages réguliers quelconques.
    - Étude fine de la structure des monoïdes de transition.
    - Premiers résultats soumis récemment (avec V. Berthé et D. Perrin).
2. Étudier les probabilités induites par les densités sur les monoïdes profinis.
    - Pont entre les aspects ergodiques et profinis des espaces symboliques.

Autres projets futurs

1. *Formalisation de la combinatoire des mots* (Isabelle/HOL), avec Š. Starosta.
2. *Semigroupes d'Ellis des mots substitutifs*, avec R. Yassawi.