

# Rapport d'activité 2023

Équipe PASTIS

## 1 Composition et domaines de recherche du groupe

Le groupe de recherche PASTIS est composé de 13 titulaires, 1 postdoc, 4 doctorants et 1 ATER (vacant) :

- Adrien Revault d'Allonnes (MCF)
- Alexandros Singh (MCF - installé en septembre 2023)
- Alice Millour (MCF)
- Anna Pappa (MCF)
- Benjamin Dupont (PRAG)
- Farès Belhadj (MCF)
- Françoise Balmas (MCF-HDR)
- Jean-Jacques Bourdin (PU)
- Louis Falissard (MCF - installé en septembre 2023)
- Nicolas Jouandeau (PU - responsable de l'équipe)
- Pablo Rauzy (MCF)
- Revekka Kyriakoglou (MCF)
- Vincent Boyer (MCF-HDR - en disponibilité)
- Sylvain Chichéry (Postdoc ANR ReComp 2022-2023)
- Jean-Pascal Palus (Doctorant 2021 - contrat doctoral ED CLI)
- Marina Seghier (Doctorante 2023 - contrat doctoral ED CLI)
- Maroua Boudabous (Doctorante 2019 - contrat CIFRE avec la société NOVAGEN)
- Stephen Obonyo (Doctorant 2021 - financement Ministère des Affaires Etrangères et de l'Europe (MEAE) / Strathmore University, Nairobi, Kenya)

Le groupe possède des compétences en synthèse d'images expressives, en enseignement de l'informatique graphique, en résolution de jeux et apprentissage automatique, en représentation du raisonnement et logique non classique, en langage naturel et apprentissage automatique, en analyse de textes, en combinatoire, en dynamique symbolique, en sécurité et privacy et en compréhension de programmes.

Jean-Pascal Palus est sous la direction de Adrien Revault d'Allonnes et Nicolas Jouandeau. Son sujet de thèse est « Dynamique et évolution de la confiance : explication et formalisation de la mésinformation ». Jean-Pascal ne s'est pas ré-inscrit en septembre 2023.

Marina Seghier est sous la direction de Alice Millour et Nicolas Jouandeau. Son sujet de thèse est « Explicabilité des modèles de TAL par les ressources : caractérisation de la variation textuelle et paradigme d'évaluation systématique ».

---

\*version 230224-1123

Maroua Boudabous est sous la direction de Anna Pappa et Françoise Balmas. Son sujet de thèse est « Modélisation d'un système intelligent d'analyse prédictive des données textuelles massives pour l'aide à la décision ».

Stephen Obonyo est sous la direction de Nicolas Jouandau. Son sujet de thèse est « Bioinformatic Computation With Tree Search and Learning ».

## 2 Bilan en chiffres

Pour l'année 2023, le groupe PASTIS compte :

- publications en conférence et revues : 9
- interventions académiques sans publication : 11
- publications grand public/médiation/valorisation : 4
- interventions grand public/médiation/valorisation : 7
- soutenances de thèse et HDR : 0
- organisations d'évènements scientifiques : 3
- participations à des comités de programme/review : 16
- projets en cours : 2
- nouveaux projets déposés : 1
- implications dans les sociétés savantes, GdR et centres de recherche : 4

Le groupe est impliqué dans le Conseil d'administration de l'AFIG<sup>1</sup>, dirige un chapitre de IEEE CIS France<sup>2</sup>, est impliqué dans le centre de recherche Geopolitique de la datasphere (GEODE)<sup>3</sup>, est impliqué dans le Conseil enseignement de EUROGRAPHICS<sup>4</sup>.

## 3 Soutenances de thèse

(aucune)

## 4 Organisations d'évènements scientifiques

1. 6ème édition du concours national API8 ([api8.fr](http://api8.fr)). **F. Belhadj**, **B. Dupont** et **A. Revault d'Allonnes**. journée de présentation des travaux. Université Paris 8, 31 mai 2023
2. organisation et animation de la track 64k du concours API8. **F. Belhadj**. 36èmes journées de l'AFIG 2023, conférence jFIG, Université de Montpellier, 8 novembre 2023
3. co-organisation de la table ronde sur le thème « Des compétences dispensées aux savoir-faire attendus : le cas de l'IG ». **F. Belhadj**. 36èmes journées de l'AFIG 2023, conférence jFIG. Université de Montpellier, 8 novembre 2023

---

1. <https://www.asso-afig.fr>

2. <http://ieee-ci.lip6.fr/>

3. <https://geode.science/>

4. <https://www.eg.org/>

## 5 Participations à des comités de programme/review

- IEEE Conference on Games (IEEE-CoG)
- Advances in Computer Games (ACG)
- 38th AAAI Conference on Artificial Intelligence (AAAI)
- 22nd IFIP International Conference on Entertainment Computing (IFIP-ICEC)
- 26th European Conference on Artificial Intelligence (ECAI)
- Revue TAL (numéro thématique “la robustesse des systèmes de traitement automatique des langues”)
- Italian philosophy journal Etica & Politica/Ethics & Politics
- The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)
- Games and NLP Workshop at LREC-COLING
- Eurographics Education program
- Special Interest Group on Computer Graphics and Interactive Techniques (ACM-SIGGRAPH)
- 32<sup>èmes</sup> rencontres francophones sur la logique floue et ses applications (LFA)
- International Conference on Neural Information Processing (ICONIP)
- TextMine - Groupe de travail sur la fouille de textes - Association Internationale Francophone d’Extraction et de Gestion des Connaissances (EGC)
- Nature Scientific Reports (Sci Rep)
- International Conference on the Theory of Information Retrieval (ICTIR)

## 6 Projets en cours

L’équipe participe a plusieurs projets, financés par Paris8 et/ou par l’ANR.

### 6.1 ANR ReComp « Research on Realtime Compliance Mechanism for AI »

**Porteur (à P8) :** Pablo Rauzy

**Partenaires :** Université Paris 8 (France), Sorbonne Université (France), Institut Fur Angewandte Informatik (Allemagne), National Institute of Informatics (Japon).

**Date de début :** 03/2021

**Date de fin :** 02/2024

**URL :** <http://research.nii.ac.jp/RECOMP/>

**Travaux réalisés en 2023 :** Poursuite des travaux de Sylvain Chichéry comme postdoc sur l’implémentation de Capacity en programmation logique (Answer Set Programming) en étroite collaboration avec l’équipe ACASA du LIP6.

### 6.2 AAP Paris 8 « Graphes, Reconnaissabilité, Apprentissage et Symétries »

**Porteurs :** Revekka Kyriakoglou et Nicolas Jouandeau

**Partenaires :** Université Paris 8 (France), Université Technique Tchèque à Prague (Tchéquie), Université Vrije à Amsterdam (Pays Bas)

**Date de début :** 01/2022

**Date de fin :** 12/2023

**URL :** <https://gras.up8.site/>

**Travaux réalisés en 2023 :** Poursuite des travaux du groupe sur la combinaison entre recherche arborescente et apprentissage par renforcement appliquée et sur la complexité. Application à la conception, à la rétro-synthèse de molécules et de protéines et à la résolution de jeux puzzle (Morpion Solitaire) et de jeux à deux joueurs (Breakthrough), et à la résolution de jeux combinatoires (Nim et Wythoff, version classique, version misère, à information complète, à information imparfaite). Participation de Revekka Kyriakoglou à la conférence Empirical Methods in Natural Language Processing (EMNLP).

## 7 Nouveaux projets déposés

### 7.1 Nouveaux projets académiques

**AAP Paris 8 « Création Lexique Innovation Crowdfunding (CLexIC) »**

**Porteuse :** Anna Pappa, en collaboration avec Alice Millour et Revekka Kyriakoglou

**Partenaires :** - Aggeliki Fotopoulou ILSP

<https://www.ilsp.gr/en/department/natural-language-processing-and-language-infrastructures/>

**Date de début :** 01/2023

**Date de fin :** 12/2023

**URL :** <https://glacon.up8.site/projects.html#clexic>

**Description :** Ce projet est une continuation du travail effectué dans le cadre du projet MALANTIN. Dans le cadre du projet CLexIC, nous voulons identifier et analyser les termes qui désignent l'innovation non technologique ainsi que le contexte dans lequel ces termes sont employés, dans écosystème du crowdfunding. Les nouveaux termes seront comparés aux termes déjà désignés par experts (pendant le projet MALANTIN) et par apprentissage sur corpus, composé des descriptions web des entreprises ainsi que de leurs bilans annuels.

**Travaux réalisés en 2023 :** Poursuite du travail de stage de Kevin Quach sur l'utilisation de l'apprentissage par transfert pour la création d'un lexique spécialisé et soumission du résumé d'un article dans JADT 2024.

### 7.2 Nouveaux projets industriels

Pas de nouveaux projet industriels pour cette année.

## 8 Activité du groupe par domaine

En synthèse d'images expressives, l'équipe initie des travaux autour de l'abstraction et de la synthèse de peinture vivante à partir de flux vidéo de plans fixes, d'une part par préfiguration du vide, et d'autre part, par l'expression fluidifiée du mouvement. Un projet a été rédigé et déposé pour les AAP 2024 de Paris 8, il permettra à court-terme de financer le besoin matériel et de mission, et à moyen terme d'aider à créer des collaborations au niveau national, dans un contexte où, actuellement, l'équipe participe au montage d'un consortium d'académiques et d'industriels intéressés par le rendu expressif. Ce consortium a pour objectif de déposer plusieurs projets dans le cadre du PEPR ICCARE.

En enseignement de l'Informatique Graphique (IG), participation aux reviews de EUROGRAPHICS Education et à celles de SIGGRAPH (external reviewer). Publication d'un article sur les meilleurs papiers d'enseignement en Informatique Graphique de 2022 [p1]. Participation active au GT-Enseignement de l'AFIG et à l'organisation et animation de la table ronde sur

l'enseignement de l'Informatique Graphique aux J.FIG 2023 à Montpellier [i1] (autant pour les travaux préparatoires, que pour l'animation de la table ronde réunissant industriels et académiques du domaine). Intervention grand public introduisant la modélisation de fractales lors du festival Maths En Ville – fête de la science, édition 2023 [m1].

L'équipe a repris l'organisation du concours annuel de synthèse d'images API8<sup>5</sup> après trois années d'absence liée au Covid. Pour cette 7ème édition, notre concours s'est ouvert à la communauté nationale d'informatique graphique en organisant la finale de la *track 64k* (session de *demomaking* avec une contrainte sur la taille des démonstrations qui ne dépassent les 64 ko) lors de la session "Concours de rendu" des J.FIG 2023 à Montpellier et c'est le public de la conférence nationale qui a pris part au vote pour le classement des trois meilleures démonstrations (plus d'une centaine de votants). De nouveaux travaux ont également été commencés en apprentissage par transfert et en meta-apprentissage de réseaux de neurones profonds, par une approche de caractérisation géométrique de leurs fonctions objectif.

En résolution des jeux et apprentissage automatique, nous avons publié un article sur la retrosynthèse moléculaire [p2]. Les tests ont été réalisés sur 5 datasets proposés par Taneda en 2012, Kleinkauf en 2015, et Runge en 2018. Les résultats permettent en moyenne de résoudre des motifs plus longs. Dans le cadre d'une collaboration avec IBM-Nairobi, nous avons contribué à la publication d'un article sur la sélection des échantillons pour la conception de modèles inverses [p3] permettant une mise en oeuvre fiable et flexible, et à la publication d'un article [p4] sur la pondération progressive d'hyperparamètres réduisant le nombre d'itérations d'apprentissage mais augmentant la convergence vers la solution. Ces travaux ont donné lieu à une présentation au Leiden Institute of Advanced Computer Science (LIACS) de l'université de Leiden [i2] et à l'université Vrije d'Amsterdam [i3]. Dans le cadre d'une collaboration avec Jean-Noël Vittaut de l'équipe LFI du LIP6, nous avons conçu le programme SuperPudim qui a obtenu la médaille de bronze aux Computer Olympiads de l'ICGA<sup>6</sup> dans la catégorie Breakthrough.

En représentation du raisonnement et logique non classique, des travaux sont poursuivis sur la modélisation de la confiance et de la croyance. La publication d'articles sur un formalisme multi-modal de la croyance en une information est à l'étude.

En traitement automatique du langage naturel (TALN), nous poursuivons nos recherches sur l'annotation de corpus, notamment en proposant un processus d'annotation semi-supervisé pour identifier et étiqueter les aspects explicites d'un corpus initialement non étiqueté [p5]. Les travaux concernant l'évaluation en TALN se poursuivent également, avec un accent donné à l'explicabilité des modèles *par les ressources*, notamment dans le cadre de la thèse de Marina Seghier, initiée en septembre 2023. Ces travaux portent sur l'apport d'une description linguistique quantitative des ressources d'évaluation pour expliquer la variation des performances observées [p6].

Une partie du projet Litte\_Bot initié en 2022 sur le recueil et le traitement de données qui ont servi pour l'apprentissage profond pour le bot de Molière, a été présenté lors de la conférence LIBER [p7]. L'expérience vis-à-vis de l'utilisation des sciences participatives dans le cadre du traitement automatique des langues peu dotées a fait l'objet d'une invitation à co-animer la table ronde d'ouverture de la conférence Contribuling 2023 [i4]. En parallèle, de nouveaux travaux cherchant à caractériser les capacités de généralisation des modèles de langues (Large Language Model) dans un cadre d'apprentissage sur corpus de petite taille ont permis de développer de nouvelles méthodes d'estimation robustes de métriques de validation durant l'ajustement paramétriquement efficace de modèles de langues [p8]. Dans le contexte d'utilisation croissante d'agents conversationnels tels que ChatGPT, un projet a été rédigé et déposé dans le cadre de l'AAP 2024 P8 afin d'étudier les risques et les opportunités pédagogiques que ces agents offrent dans le cadre de l'enseignement de l'informatique.

---

5. <https://www.api8.fr/>

6. <https://icga.org/>

En combinatoire [i5], nos recherches se sont concentrées sur les aspects bijectifs, analytiques et probabilistes de la combinatoire des cartes (de graphes sur des surfaces) et des objets connexes. Ceci a notamment donné lieu à un exposé invité lors d’une journée thématique intitulée “Elimination for Functional Equations” dans le cadre de la semaine “Computer Algebra for Functional Equations in Combinatorics and Physics” du trimestre thématique “Recent Trends in Computer Algebra” de l’Institut Henri Poincaré. Dans le cadre de ce travail, une proposition de projet AAP Paris 8 a été soumise pour explorer les aspects algébriques de la combinatoire des cartes en relation avec la combinatoire du groupe modulaire. Le projet implique de nouvelles collaborations avec des membres du Département de Mathématiques de l’Université de l’Egée, qui est également membre de l’alliance ERUA à laquelle Paris 8 participe.

En dynamique symbolique et combinatoire des mots, notre recherche se concentre sur la notion de la *reconnaissabilité* et la *circularité* des mots dans les langages substitutifs. Le terme reconnaissabilité fait référence à la décomposition unique des mots et à l’identification de leurs pré-images (antécédents). Nous travaillons actuellement dans une étude approfondie de ces définitions, en examinant leur équivalence et les différences entre les différentes classes de morphismes.

En sécurité et privacy, l’essentiel de l’activité de l’équipe pour l’année 2023 se concentre autour de la compréhension et de l’explication des limites techniques des *blockchains*. Cette activité est menée de façon transdisciplinaire, aussi bien académiquement qu’auprès du grand public et des institutions. Ces travaux ont donc donné lieu à une publication académique dans la revue multidisciplinaire *Terminal* [p9], à une série de conférences en séminaires et colloques universitaires [i6, i7, i8, i9], d’interventions auprès d’institutions (ministère des finances [i10], ministère de la justice [i11]), ainsi qu’à de nombreuses interventions auprès du grand public, à travers des articles de journaux [a1] et de blogs [a2, a3, a4], lors d’événements particuliers (théâtre [m2], salons et festivals [m3, m4, m5, m6]), ou encore à l’invitation d’associations de médiation culturelle et scientifique [m7]. Un projet a été rédigé et déposé pour les AAP 2024 de Paris 8, dans l’objectif d’organiser un symposium sur la question des limites des blockchains, en y faisant intervenir des chercheur-es de différentes disciplines (informatique, géopolitique, économie) ainsi que des professionnel-les de la médiation et du journalisme scientifique afin de rendre accessible l’évènement au grand public.

## 9 Présentations et réunions de travail réalisées dans le cadre du séminaire SOIF

Dans la continuité de l’année 2022, le Séminaire d’Ouverture aux Informatiques et de Formation (SOIF) a été maintenu afin d’échanger sur les travaux en cours des membres de l’équipe. La fréquence des séminaires est passée à une fois par semaine en septembre 2023 afin d’avoir plus d’échanges dans des réunions de travail.

**Alice Millour** (maîtresse de conférences dans l’équipe PASTIS)

**Date** : 26 janvier 2023

**Titre** : Katana and Grand Guru : the Game of Lost Words (demo)

**Résumé** : We present a prototype of a role playing game which allows to both 1- crowdsource lexical units (including idioms) for a language and 2- help the player improve their knowledge of the language. Our implementation of the game is focused on non-standardized languages, for which the intergenerational transmission is not as efficient as it used to be. In order to address this, we incentivize the participation of a “Grand Guru”, from whom the player needs help to fulfill their mission.

**Cyrielle Ndongo Ngondi** (étudiante en troisième année de licence informatique à l'université Paris 8)

**Date** : 9 février 2023

**Titre** : Présentation et démonstration de "Box Area"

**Résumé** : Présentation des travaux de conception et démonstration du jeu "Box Area", développé dans le cadre de mon séjour à Spelkollektivet, en Suède, financé par la Mairie de Saint-Denis et l'UFR STN (département PIF). Description du jeu : « Votre patron n'est pas au bureau. Tous le monde s'ennuie. Qu'est ce que vous pourriez faire ? Une course dans des boîtes avec vos collègues ! ».

**Cherifa Ben Khelil** (post-doc à l'Université de Tours dans le projet AAC4All)

**Date** : 23 mars 2023

**Titre** : Améliorer les modèles de prédiction de mots pour la Communication Améliorée et Alternative chez les enfants

**Résumé** : La Communication Améliorée et Alternative (CAA) vise à suppléer ou remplacer la parole et/ou l'écriture pour les personnes souffrant de troubles graves de la parole et du langage. Cela concerne une grande diversité de handicaps pour lesquels il existe une gamme tout aussi large de réponses en termes d'aides techniques. Un système typique de CAA consiste en un clavier virtuel à partir duquel le texte est saisi à l'aide d'un dispositif de contrôle dédié, ainsi qu'une synthèse vocale. Quel que soit le système considéré, la saisie de texte est lente et fatigante. Pour permettre une saisie plus rapide, deux approches complémentaires peuvent être envisagées : la sélection rapide des touches grâce à une conception optimisée des claviers, et l'économie de frappe grâce aux techniques de prédiction de mots. Les mots prédits sont affichés dans une liste de prédictions, permettant ainsi la complétion de mots et une saisie de texte plus rapide. Les modèles de prédiction reposent sur des techniques d'apprentissage automatique. Ils sont entraînés sur de grands corpus extraits de ressources web ou de journaux et sont censés être représentatifs de la langue de communication. Cette approche soulève la question de l'adaptation de la prédiction à tous types d'utilisateurs. Au cours du projet AAC4All (Alternative & Augmentative Communication designed for all), nous chercherons à développer des moteurs de prédictions adaptées au langage enfantin, et plus globalement qui s'adaptent à l'évolution des compétences langagières de l'utilisateur. Nous avons construit un nouveau corpus francophone dédié au langage enfantin. Cette ressource a permis l'entraînement de modèles de prédiction spécifiques qui répondent mieux aux besoins des jeunes utilisateurs. Des expériences menées sur des textes réels montrent les avantages de cette adaptation en termes d'économie de frappe. En particulier, l'amélioration observée de l'économie de frappe permet de limiter le nombre de mots prédits à afficher. Cela ouvre des opportunités pour reconsidérer l'organisation du clavier virtuel ainsi que les styles d'interaction qu'il offre.

**Mariya Evtimova-Gardair** (ATER en informatique à l'Université Paris 1 Panthéon-Sorbonne)

**Date** : 30 mars 2023

**Titre** : Plate-forme multilingue basée sur la connaissance pour les étudiants étrangers en tant que guide culturel et éducatif

**Résumé** : Cette année, le début de l'invasion russe en Ukraine a entraîné une augmentation du nombre d'étudiants ukrainiens dans les systèmes éducatifs français. Les statistiques de l'UNESCO montrent que 30 000 ukrainiens sont entrés en France le 25 mars de cette année et que 6 873 étudiants ukrainiens ont été inscrits dans le système éducatif français. Ces statistiques sont également valables pour les autres pays de l'UE. Cette vague d'étudiants étrangers a conduit à la nécessité de créer un outil qui facilitera leur intégration. Cet outil vise à faciliter le processus d'intégration des étudiants étrangers, en particulier lors de leur première année d'études

à l'étranger. Cet outil numérique créera une plateforme multilingue pour un guide culturel et éducatif avec des informations pour les étudiants étrangers. Pour mettre en œuvre le projet et fournir des informations de grande valeur aux étudiants, il est nécessaire de transformer les données en connaissances multilingues en utilisant les technologies sémantiques et leur application aux algorithmes de recherche de correspondance.

**Mehdi Naima** (post-doctorant à RWTH, Aachen, en Allemagne)

**Date** : 6 avril 2023

**Titre** : Extending Brandes algorithm to improve betweenness centrality computation in temporal graphs with discrete and continuous time

**Résumé** : Betweenness centrality assesses the importance of nodes in a graph, it has been used for example in social, biological and research collaboration networks. Moreover, betweenness centrality has been used in graph partitioning and community detection in the well-known Girvan-Newman algorithm. This centrality measure is based on the enumeration of shortest paths passing through a node. A simple approach to compute betweenness centrality for all the nodes of a static graph is to use Floyd-Warshall algorithm that runs in  $O(n^3)$ . Brandes in 2001 published an algorithm that runs in  $O(nm+n^2\log n)$  on weighted graphs, it is still considered one of the best theoretical results on the question. Betweenness centrality has also been extended to temporal graphs. Temporal graphs have edges that bear labels according to the time of the interactions between the nodes. Betweenness centrality has been extended to the temporal graph settings, and the notion of paths has been extended to temporal paths. We will see that we are able to deploy Brandes algorithm to its full extent and improve the running time of recent results to  $O(nmT + n^2T\log(nT))$ . We will also discuss how Brandes algorithm can also be generalized to stream graphs which are dynamic graphs with continuous time and dynamicity on the nodes.

**Alexandros Singh** (docteur de l'Université Sorbonne Paris Nord)

**Date** : 11 avril 2023

**Titre** : La combinatoire du lambda-calcul et des cartes : bijections et propriétés limites

**Résumé** : Cet exposé concerne la combinatoire des cartes et leur interaction avec le lambda calcul linéaire. En utilisant une combinaison de nouvelles techniques bijectives et analytiques, nous explorerons la structure de grandes cartes cubiques aléatoires et de lambda-termes linéaires, en nous concentrant sur l'étude de diverses statistiques sur de tels objets, en particulier celles liées à la présence de divers motifs et leur connexion avec la dynamique de la bêta-réduction sur les termes linéaires.

**Reda Belaiche** (ATER à l'Université Paris Cité)

**Date** : 13 avril 2023

**Titre** : Analyse temps réel des micro-expressions par vision artificielle

**Résumé** : Les technologies de l'interaction homme machine se concentrent de plus en plus sur l'être humain, que ce soit sur son identité, ou bien sur son état physique et mental. Des progrès conséquents ont été réalisés depuis quelques décennies. Cependant l'étude des pensées et des émotions reste encore un domaine peu développé, mais qui a commencé à grandement gagner en intérêt. Dans ce domaine, l'analyse des expressions faciales est le traitement préférentiel. Contrairement aux macro-expressions, visible à l'œil, les micro-expressions sont un type d'expressions faciales involontaires, extrêmement rapides et de très faible intensité. La communauté scientifique en vision par ordinateur étudie depuis quelques années les façons de reconnaître automatiquement les micro-expressions à l'aide de cameras rapides et de programmes informatiques. Il s'agit néanmoins d'un problème difficile, de par la nature de ces micro-expressions. À partir des dernières avancées techniques en machine learning, nous avons proposé plusieurs méthodes de



reconnaissance des micro-expressions basées sur la vision tout en mettant l'accent sur la rapidité d'exécution et les faibles besoins en mémoire.

**Karim Tabia** (maître de conférences HDR, dans l'équipe CRIL de l'Université d'Artois)

**Date** : 14 avril 2023

**Titre** : Raisonnement sous incohérence : Application en gestion de contenus vidéos et en IA explicable

**Résumé** : L'incertitude et l'incohérence sont deux aspects omniprésents des croyances, des connaissances et des données dans de nombreux domaines. Par conséquent, le raisonnement à partir d'informations incohérentes, généralement représentées sous forme de formules logiques, est un problème important en intelligence artificielle. Cette présentation est consacrée à quelques-unes de nos contributions au raisonnement avec des informations incohérentes illustrées sur deux applications : la gestion de contenus vidéos et l'IA explicable. La première, abordée dans le cadre du projet européen H2020-MSCA-RISE-2015 AniAge, concerne l'interrogation de bases de données hétérogènes et massives avec des parties assertionnelles affectées par l'incertitude et éventuellement des conflits. La deuxième application traite de l'explication des modèles d'apprentissage automatique en réduisant la recherche ou l'énumération d'explications à des problèmes bien connus de calcul d'ensembles de conflits et d'ensembles de corrections et pour lesquels des outils génériques existent.

**Marina Seghier** (doctorante dans l'équipe PASTIS)

**Date** : 19 octobre 2023

**Titre** : Caractérisation objective des catégories textuelles pour le TAL : classification non-supervisée basée sur des descripteurs linguistiques

**Résumé** : Les systèmes développés et de plus en plus répandus aujourd'hui, sont présentés comme étant très performant pour un grand nombre de tâches. Or, les performances annoncées ne sont pas toujours celles rencontrées selon les types de ressources textuelles auxquels ces outils sont confrontés. En effet, on peut par exemple observer un différentiel de performances important entre différents types de texte pour la tâche de reconnaissance d'entités nommées (Millour et al., 2022) et d'étiquetage morpho-syntaxique. Cependant, les typologies textuelles existantes, fondées sur une classification des catégories *a priori* sans justification linguistique, ne permettent pas d'expliquer ce différentiel. Ce travail s'inspire des travaux de Biber, 1988 et Passonneau et al., 2014, sur la caractérisation de la variation textuelle à partir de descripteurs linguistiques. Cette étude a permis de retrouver certaines similitudes avec les dimensions des travaux précédents, et ouvre la voie à de nombreuses extensions.

**Louis Falissard** (maître de conférences dans l'équipe PASTIS)

**Date** : 26 octobre 2023

**Titre** : Gros modèles de langue, sous-espaces et méthodes paramétriquement efficaces

**Résumé** : L'avènement au cours des dernières années des gros modèles de langues a été la source d'une évolution considérable des applications de méthodes d'apprentissage profond en traitement automatique des langues. Ces modèles, pré-entraînés de manière non-supervisée sur des corpus de données textuelles massifs, permettent notamment l'ajustement de puissants modèles neuronaux à partir de quelques milliers, voire centaines d'observations, avec des performances de généralisation qui demandaient encore il y a quelques années plusieurs millions d'exemples. Les méthodes dites "paramétriquement efficaces", notamment, permettent d'ajuster ces modèles de langues en réduisant significativement leurs besoins en terme de puissance de calculs, tout en affichant des performances typiquement meilleures qu'un ajustement classique sur de petits jeux de données. On s'intéressera dans ce séminaire à présenter ces méthodes, ainsi que leur

utilisation dans le cadre de la méthode des “sous espaces”, pour améliorer les capacités de généralisation et de transfert de ces modèles dans un cadre où très peu de données sont disponible pour l’entraînement du modèle (entre 50 et 500 observations).

**Alexandros Singh** (maître de conférences dans l’équipe PASTIS)

**Date** : 16 novembre 2023

**Titre** : Quelques équations fonctionnelles pour des cartes

**Résumé** : Le but de cet exposé exploratoire est de présenter diverses équations fonctionnelles relatives à l’énumération des cartes. En particulier, nous nous intéresserons à des questions telles que : comment ces équations sont-elles reliées et peuvent-elles être inter-dérivées par des méthodes de combinatoire ou d’algèbre différentielle ? Cette présentation s’appuie sur des travaux en cours avec Olivier Bodini et Konstantinos Tsagkaris.

**Philippe Guillot** (ex-maître de conférences en mathématiques à Paris 8)

**Date** : 30 novembre 2023

**Titre** : Le Vlisp, Le Lisp de Paris 8

**Résumé** : Le département informatique de l’université de Vincennes a développé dans les années 1970 toute une famille d’interprètes LISP, appelés VLISP, V comme Vincennes, sous l’impulsion de Patrick Greussay, Harald Wertz, Françoise Balmas, et Daniel Goosens notamment. Ce travail a ensuite été poursuivi à l’INRIA par Jérôme Chailloux avec un produit commercial : Le\_Lisp. Le projet Le Vlisp, débuté à l’occasion du confinement de 2020, est pensé dans cette continuité. Il présente aujourd’hui un état d’avancement suffisant pour être présenté au département d’informatique. Nous aborderons dans cette présentation : ce qu’il est possible de faire avec, son architecture générale, son environnement de programmation, et les caractéristiques du langage. En perspective seront évoqués les projets possibles et les travaux à poursuivre concernant le développement et l’utilisation du langage.

## 10 Productions scientifiques pour l’année 2023

### 10.1 Publications en revues et conférences

*Une astérisque marque l’auteur·ice ayant présenté les travaux lors de la conférence.*

- [p1] E. Paquette and **J.-J. Bourdin**. Foreword to the special section on best papers of the eurographics 2022 education papers program. *Computers & Graphics*, 112 :A6–A7, 2023.
- [p2] **S. Obonyo\***, **N. Jouandeu** and D. Owuor. RetroG : Retrosynthetic Planning with Tree Search and Graph Learning. In *Machine Learning for Drug Discovery Workshop at International Conference on Learning Representations (ICLR–MLDD)*, pages 1–12, 2023.
- [p3] T. Rumbell\*, C. Wanjiru, I.O. Mulang, **S. Obonyo**, J. Kozloski, and V. Gurev. Sequential data-consistent model inversion. In *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*, pages 1–4, 2023.
- [p4] **S. Obonyo\***, I.O. Mulang, T. Rumbell, C. Wanjiru, and V. Gurev. Feature Importance of Random Search for Hyperparameter Optimization of Data-Consistent Model Inversion. In *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*, pages 1–4, 2023.
- [p5] **M. Boudabous\*** and **A. Pappa**. Explicit aspect annotation via transfer and active learning. *Procedia Computer Science*, 225 :1124–1133, 2023. 27th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES 2023).

- [p6] **M. Seghier\***, **A. Millour** and J.-Y. Antoine. Descripteurs Linguistiques et Caractérisation Objective des Catégories Textuelles. In Karèn Fort, Claire Gardent, and Yannick Parmentier (Éds.), editors, *5èmes journées du Groupement de Recherche CNRS “Linguistique Informatique, Formelle et de Terrain”*, pages 106–112, Nancy, France, November 2023. GdR LIFT, CNRS.
- [p7] **A. Pappa\***, S. Szoniecky, B. Berenguer, J. Huthwohl, C. Quach and A. Laborderie. Designing a database for a classical theatrical chatbot. In *52nd LIBER Annual Conference*, pages 153–155, 2023.
- [p8] **L. Falissard\***, V. Guigue and L. Soulier. Improving generalization in large langue model by learning prefix subspaces. In *Findings of the Association for Computational Linguistics : Empirical Methods in Natural Language Processing (EMNLP)*, pages 11474–11483, 2023.
- [p9] **P. Rauzy**. Promesses et (dés)illusions : une introduction technocritique aux blockchains. In *Terminal, "Blockchains, quels enjeux ?"*, volume 136, pages 42–58, 2023.

## 10.2 Interventions académiques sans publication

- [i1] **F. Belhadj**. Des compétences dispensées aux savoir-faire attendus : le cas de l'IG. Animation de la table ronde de la session enseignement des journées françaises d'informatique graphique J.FIG'23, à Montpellier. 8 novembre 2023.
- [i2] **S. Obonyo**. Computational biology with tree search and learning. Invited talk, Leiden Institute of Advanced Computer Science, Leiden University. 14 décembre 2023.
- [i3] **S. Obonyo**. RNA inverse folding with langage model and tree search. Invited talk, Vrije Universiteit Amsterdam. 16 décembre 2023.
- [i4] **A. Millour**. Table ronde d'ouverture de Contribuling 2023. Conférencière invitée, Paris, France. 12 mai 2023.
- [i5] **A. Singh**. On some functional equations for maps. Topical day "Elimination for Functional Equations", Computer Algebra for Functional Equations in Combinatorics and Physics, Institut Henri Poincaré (IHP). 11 décembre 2023.
- [i6] **P. Rauzy**. Introduction technocritique aux blockchains. Conférence invitée, séminaire du laboratoire CITI, Lyon. 10 février 2023.
- [i7] **P. Rauzy**. Introduction technocritique aux blockchains. Conférence invitée, Faculté de Droit, économie, gestion de l'Université d'Angers. 14 avril 2023.
- [i8] **P. Rauzy**. Introduction technocritique aux blockchains. Conférence invitée, Workshop *Chaînes de blocs et monnaie : regards croisés*, Grenoble. 4 mai 2023.
- [i9] **P. Rauzy**. La vérité sur la blockchain. Conférence invitée, séminaire du laboratoire MIA INRAE / AgroParisTech / Université Paris-Saclay, Palaiseau. 14 décembre février 2023.
- [i10] **P. Rauzy**. Audition par une mission de l'Inspection Générale des Finances et du Ministère de l'Industrie sur les blockchains et crypto-actifs. Ministère des finances, Paris Bercy. 12 janvier 2023.
- [i11] **P. Rauzy**. Entretien avec Antoine Meissonnier (chef du département des archives, de la documentation et du patrimoine du Ministère de la Justice) au sujet des blockchains et de l'archivage. Paris. 10 juillet 2023.

## 10.3 Publications grand public / médiation / valorisation

- [a1] **P. Rauzy**. Réseaux sociaux : comparaison tout à fait partielle entre Twitter et Mastodon. *Alternative libertaire*. n°334, janvier 2023.

- [a2] **P. Rauzy**. Analyse de vocabulaire technique : “minage”. Article de blog. 8 mai 2023.
- [a3] **P. Rauzy**. Compte-rendu de lecture : *No Crypto* de Nastasia Hadjadji. Article de blog. 23 mai 2023.
- [a4] **P. Rauzy**. Compte-rendu de lecture : *Cryptocommunisme* de Mark Alizart. Article de blog. 21 juin 2023.

#### 10.4 Interventions grand public / médiation / valorisation

- [m1] **F. Belhadj**. Comment créer un paysage virtuel avec des fractales ? Conférence et rencontre lors du festival Maths en Ville, fête de la science, édition 2023, compagnie Terraquée. 10 octobre 2023.
- [m2] **P. Rauzy**. Rencontres autour des blockchains et crypto-actifs suite aux représentations de la pièce *HPNS, Marché pirate sur le darknet* de Léonard Matton. Théâtre de la Reine Blanche, Paris. 16, 17 et 18 mars 2023.
- [m3] **P. Rauzy**. Introduction technocritique aux blockchains. Conférence aux Journées du Logiciels Libres, Lyon. 1er avril 2023.
- [m4] **P. Rauzy**. Introduction technocritique aux blockchains. Conférence au festival Pas Sage En Seine, Choisy-le-Roi. 16 juin 2023.
- [m5] **P. Rauzy**. La vérité sur le blockchain. Conférence au festival Capitole du Libre, EN-SEEIHT, Toulouse. 18 novembre 2023.
- [m6] **P. Rauzy**. La vérité sur le blockchain. Conférence au salon Open Source Experience, Palais des congrès, Paris. 6 décembre 2023.
- [m7] **P. Rauzy**. La vérité sur le blockchain. Conférence invitée par les Petits Débrouillards Occitanie à l’Eurêkafé, Toulouse. 16 novembre 2023.