Novembre
2022

numéro
20

1024 Bulletin de la Société informatique de France



COMITÉ DE RÉDACTION

SYLVIE ALAYRANGUES
Université de Poitiers

OLIVIER BAUDON
Université de Bordeaux

YVES BERTRAND
Université de Poitiers

JEAN-PAUL DELAHAYE
Université de Lille

ISABELLE DEBLED-RENNESSON
Université de Lorraine

GIUSEPPE DI MOLFETTA
Université Aix-Marseille

MARIE DUFLOT-KREMER
Université de Lorraine

THOMAS FERNIQUE
CNRS, Université Paris 13

FRÉDÉRIC HAVET
CNRS, Université Côte d'Azur

PHILIPPE MARQUET
Université de Lille

PATRICE NAUDIN
Université de Poitiers

PIERRE PARADINAS
CNAM Paris

NICOLAS PASSAT
Université de Reims Champagne-Ardenne

MICHEL RAYNAL
Université de Rennes

LAURENT RÉVEILLÈRE
Université de Bordeaux

NATHALIE REVOL
Inria, Université de Lyon

NANCY RODRIGUEZ
Université de Montpellier

FLORENCE SÈDES
Université de Toulouse

DENIS PALLEZ, *Université Côte d'Azur, rédacteur en chef*

Contact : 1024@societe-informatique-de-france.fr

Cette œuvre est mise à disposition sous licence Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0. Pour voir une copie de cette licence, visitez http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr *ou écrivez à Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.*

SOCIÉTÉ INFORMATIQUE DE FRANCE
Institut Henri Poincaré, 11 rue Pierre et Marie Curie, 75231 Paris Cedex 05

Prix public : 32 € (adhérents SIF : −30%)
Directeur de la publication : Yves Bertrand ISSN : 2270-1419

Couverture réalisée par l'agence graindecel, https://www.bauwensceline.com.

SOMMAIRE DU Nº 20

Éditorial, *D. Pallez* 3

LES 10 ANS DE LA SIF

10 ans déjà, 10 ans seulement... *Yves Bertrand* 5
Création de la SIF, *Colin de la Higuera* 9
Table ronde des anciens présidents, *Fabien Tarissan* 21
Une association reconnue d'utilité publique , *Élisabeth Murisasco* 31
Table ronde sur la recherche en informatique, *Serge Abiteboul* 39
Discours de présentation du blog Binaire, *Serge Abiteboul* 45
Mille vingt-quatre, *Denis Pallez et Éric Sopena* 51

CODES SOURCES

Prolog, 50 ans d'histoire de l'intelligence artificielle, *Laurent Cervoni et Julien Brasseur* 55
Algol 60, 60 ans après, *Thierry Dumont* 65
Catala : formaliser la loi grâce à un langage de programmation dédié, *Denis Merigoux et Liane Huttner* 77
Un langage pour enseigner la programmation, Scheme ou Python ? *Laurent Bloch* 85

ENSEIGNEMENT

(R)évolutions de l'informatique dans le sup : orientation et pratiques pédagogiques ! *Bruno Mermet* 97
Accompagner l'enseignement de l'informatique avec Caseine, *Équipe Caseine* 107

FICTION

La ronde des avatars, *Serge Abiteboul* 123
Mon rosbif est biaisé, *Serge Abiteboul et Julia Stoyanovich* 125
Cours toujours, *Fabien Gandon* 129

HOMMAGE

Hommage à Gérard Comyn, *Pierre Paradinas, Max Dauchet* 139
Hommage à Gilles Muller, *Ben Mokhtar, Bromberg, Lawall, Réveillère et Rouvoy* 141

INFORMATIQUE ET SOCIÉTÉ

Authentification biométrique et respect de la vie privée : points de vue et limitations, *Estelle Cherrier-Pawlowski* 145

Journée « Raconter la science en temps de crise », *Journée Sciences et médias 2022* 161

INFORMATIQUE ET SES INTERACTIONS

Les réseaux d'automates booléens au cœur du calcul naturel, *Kévin Perrot et Sylvain Sené* . . 171

NOTES DE LECTURE

The Science of Science, *lu par Michel Raynal* 183

PROJETS SCIENTIFIQUES

Projet ANR (2020-2024) « Hérelles » : un cadre collaboratif unifié pour l'analyse interactive de données temporelles, *Gançarski, Cornuejols, Cremilleux, Dao, Teisseire* 185

RÉCRÉATION

Le problème des trois cartes, *Jean-Paul Delahaye* 197

Éditorial

Chères adhérentes, chers adhérents, chères lectrices, chers lecteurs,

Joyeux anniversaire !

La SIF fête ses 10 ans d'existence et, dans ce numéro, nous avons retranscrit quelques uns des discours prononcés à l'occasion du congrès anniversaire de notre association en juin 2022. Si vous avez eu la chance de venir souffler avec nous les 10 bougies de la SIF à Paris, vous avez pu découvrir, en avant première, notre nouveau logo.

Et aujourd'hui, c'est au tour de notre bulletin de se mettre à la page avec un tout nouveau visuel. Pour l'instant, seule la couverture a fait peau neuve mais je ne désespère pas d'avoir une nouvelle présentation de son contenu dans un futur proche, le plus proche possible... nous y travaillons.

Pour fêter cet anniversaire et — en plus de l'exemplaire de la bande-dessinée des « Décodeuses du numérique » envoyé avec le numéro 19 d'avril — la SIF vous offre un exemplaire papier du livre « Les entretiens de Binaire » qui a également été distribué aux participants lors du congrès.

Pour marquer encore un peu plus cet événement, et sous l'impulsion de Sylvie Alayrangues, membre du comité de rédaction, nous avons volontairement introduit plusieurs nouveautés rédactionnelles dans ce numéro. Tout d'abord, la retranscription de plusieurs discours prononcés au congrès illustrés de manière originale par Catherine Créhange [1], graphiste captant de manière humoristique l'ambiance qui s'est dégagée lors des deux jours de commémoration. Autre nouveauté : une bande-dessinée, qui raconte les circonstances dans lesquelles notre association a été reconnue d'utilité publique. Une nouvelle rubrique voit le jour : « Fiction », qui est évidemment ouverte à tous les auteurs et autrices ; vous connaissez sans aucun doute

1. https://undessinparjour.wordpress.com.

DOI:10.48556/SIF.1024.20.3

ceux qui inaugurent cette rubrique de fictions comme auteurs de nombreux articles scientifiques mais vous ne leur connaissiez peut-être pas cette aptitude à concevoir et écrire des nouvelles, aussi rafraîchissantes.

Une autre nouvelle rubrique fait également son apparition mettant en lumière la programmation et le codage informatique : elle porte le même nom « Codes sources [2] » que le séminaire dont elle est issue.

Mais il est temps maintenant de vous laisser redécouvrir toutes les rubriques et articles qui font de ce bulletin ce qu'il est. Un grand merci, une fois de plus, aux auteurs des articles qui nous ont contactés.

DENIS PALLEZ
Rédacteur en chef
1024@societe-informatique-de-france.fr

2. `https://codesource.hypotheses.org`.

10 ans déjà, 10 ans seulement...

Yves Bertrand [1]

Début 2021, la SIF bruissait déjà d'un événement qui promettait d'être marquant : elle fêterait ses 10 ans en 2022 ! Au moment où vous lisez ces lignes, c'est chose faite : la SIF a célébré ses 10 ans lors du congrès de 2022 au CNAM, trouvant un juste équilibre entre les regards tournés vers le passé et ceux esquissant les contours d'une « boussole » pour guider la SIF pour les dix ans à venir.

J'attendais beaucoup du discours de mes prédécesseurs à ce congrès pour éclairer ma réflexion de leur expérience. La richesse de leur table ronde a totalement répondu à cette attente. C'est pourquoi mon parti pris est de centrer ici mon propos sur le leur, même si le fait que la SIF soit une société savante visible et reconnue dans le paysage académique national doit énormément aux équipes et qui les ont entourés et à plusieurs personnalités du conseil scientifique et du conseil d'administration de la SIF.

Dans les propos liminaires de Colin [2] sur la naissance de la SIF, puis dans les interventions de Colin, Jean-Marc [2] et Pierre [2], j'identifie et fais miens sept champs d'action pour la SIF : les voici, pour en débattre...

Développer les actions de médiation scientifique...

En allant au contact du grand public, des parents d'élèves, des enseignants, des enfants pour démystifier l'informatique, pour qu'« algorithme », « intelligence artificielle », par exemple, ne riment plus avec « magie ». L'enjeu est de taille ! L'informatique demeure en effet abstruse aux yeux du plus grand nombre, et il est autrement plus difficile de parler d'informatique à un béotien où à un ministre que d'ergoter entre spécialistes sur les acceptions comparées de termes techniques voisins. Les

1. Président de la Société informatique de France, professeur des universités, université de Poitiers.
2. Colin de la Higuera, Jean-Marc Petit, Pierre Paradinas : précédents présidents de la SIF.

DOI:10.48556/SIF.1024.20.5

vidéos qui désacralisent les quatre piliers de l'informatique (algorithmes, langages, machines, données) conçues avec Fred Courant [3] et son équipe de « L'esprit sorcier » en 2022 sont un exemple de ce dialogue à développer avec le grand public.

Fédérer la communauté autour de l'enseignement

Parce que c'est la qualité de la formation en informatique, en particulier dans le primaire et le secondaire, qui fera, ou pas, de notre pays une nation à la pointe du numérique. Grâce au combat des pionniers depuis près d'un demi-siècle, l'informatique s'enorgueillit depuis peu de quelques lettres de noblesse : un CAPES NSI, une agrégation informatique, des classes préparatoires intégrant l'informatique, et une spécialité NSI au lycée. Mais ces avancées sont aussi fragiles qu'elles sont encourageantes : le CAPES est passé de 60 à 50 places entre 2020 et 2021 ; l'agrégation de 2021 s'est vue dotée de 20 postes ; la moitié des lycées n'offrent pas la spécialité NSI. Tout reste donc à construire pour que l'informatique puisse un jour être, comme les mathématiques, la physique, la chimie ou la biologie un constituant séculaire de la culture scientifique de chacun.

Promouvoir le duo action politique / communiqué

Évoqué par Colin, ce duo est plus d'actualité que jamais. Un exemple : l'enseignement SNT en seconde. Sous la houlette d'Isabelle [4], un communiqué prônant l'évolution de cet enseignement vers le cœur de l'informatique est associé à des démarches auprès du MEN. Réciproquement, des contacts institutionnels doivent donner lieu à des communiqués de la SIF. Avec un risque évident : celui d'essuyer d'autant plus de critiques que nos prises de positions seront nettes, fortes et clivantes. Mon credo est qu'il vaut mieux être critiqué pour des participations audacieuses au débat public que pour l'absence de la SIF dudit débat.

Mettre les individualités au service du collectif

C'est une valeur constitutive de la SIF qu'il convient de préserver à tout prix. Aujourd'hui, les individualités les plus reconnues sont au service d'une cause commune : le développement d'une informatique au service de l'humain. Comme Jean-Marc, j'en cite deux figures de proue évidentes : Serge et Gilles [5] qui mettent leur énergie, leur hauteur de vue et leur expérience au service de notre collectif. La boussole de la SIF en sera d'autant plus pertinente, lisible, pérenne et préservée des contingences politiciennes court-termistes.

Faire œuvre d'histoire et de mémoire

Dans son évolution, l'informatique est à un moment idéal pour se doter de trois mémoires. *La mémoire du matériel* : Pierre a développé l'idée d'un musée de l'informatique, central ou en réseau. Sous son impulsion, et avec le CNAM et INRIA,

3. Journaliste, grand reporter et animateur de télévision français.
4. Isabelle Debled-Rennesson, vice-présidente Enseignement de la SIF.
5. Serge Abiteboul et Gilles Dowek, anciens présidents du conseil scientifique de la SIF.

la SIF est partie prenante de la sauvegarde d'une collection matérielle conséquente. Gageons que cette action préfigure la naissance d'un musée national porté par les institutions dont l'objet est le développement de la science informatique. *La mémoire des logiciels* : saluons l'œuvre de Roberto [6] dont le projet Software heritage a déjà collecté plus de 10 milliards d'objets logiciels, et vise ni plus ni moins qu'à être une « bibliothèque d'Alexandrie » de notre discipline qui compte aujourd'hui plus de langages que l'humanité ne compte de langues et dialectes. *La mémoire des idées*. Puisse l'impulsion de membres illustres de la SIF contribuer à faire voir le jour dans notre bulletin 1024 à une histoire des idées de l'informatique racontée aux générations futures par celles et ceux qui les ont développées depuis plus d'un demi-siècle.

Professionnaliser la relation de la SIF avec le monde socio-économique

Notre congrès 2022 m'a convaincu que notre association est née et demeurera société savante académique. Il serait vain de viser une quelconque parité interne avec des membres industriels qui se passent fort bien de notre concours direct, tellement leurs enjeux financiers et technologiques leur semblent éloignés des nôtres. En revanche, les puissantes associations qui les représentent, comme le CIGREF et Numeum, ont manifesté à maintes reprises leur ouverture à nos positions concernant la médiation, la formation, la parité femme-homme, la formation en mathématiques et en sciences, le développement durable, la responsabilité sociétale. Lisez leurs tribunes, leurs propositions : vous serez surpris de constater à quel point leurs préoccupations sur ces thématiques rejoignent les nôtres. C'est donc avec ces associations qu'il convient de tisser des liens durables.

L'inégalité femme-homme

Sujet difficile s'il en est, il fut abordé rapidement par mes prédécesseurs. La part des femmes dans notre discipline demeure à un niveau très bas depuis trente ans. La part des filles au lycée en informatique stagne autour de 13 % en terminale. En 2021, seules 2 215 filles ont choisi la spécialité NSI en terminale. L'industrie du numérique demeure très largement masculine. Jusqu'en 2019, le lycée contenait les disparités femmes – hommes dans les disciplines scientifiques se faisant jour dès le CE1. Depuis trois ans, ces disparités reprennent de la vigueur à un rythme préoccupant pour l'avenir de notre discipline dans le monde socio-économique. Agissons de conserve avec ses principaux représentants pour infléchir cette inquiétante trajectoire. Agissons aussi avec eux pour proposer des journées « informatique et société » comme celle du 4 octobre 2022 à Bordeaux sur les « Algorithmes d'aide à la décision publique » et comme celle du 22 novembre 2022 à Paris sur les « Infrastructures pour la souveraineté numérique » en donnant aux femmes systématiquement la place qu'elles méritent : la même que celles des hommes. Eprouvons notre logique interne jusqu'à faire en sorte que le prochain président de la SIF soit... une présidente ?

6. Roberto di Cosmo, `https://www.dicosmo.org`.

Les chantiers sont nombreux ; les enjeux, majeurs. Ils sont à la hauteur de l'omniprésence de l'informatique dans notre société. Oeuvrons donc, ensemble, en privilégiant l'action externe aux débats internes, à rendre la SIF omniprésente dans les débats qui animent les sphères académique, institutionnelle et sociétale dans lesquelles elle évolue.

Création de la SIF

Colin de la Higuera [1]

Le 31 mai 2012, l'assemblée générale extraordinaire des personnels enseignants et chercheurs en informatique de France (SPECIF [2]) décidait de sa transformation en Société informatique de France avec une modification substantielle de ses statuts, de ses objectifs, de ses missions. Plus précisément, il y eu 131 voix pour la transformation de SPECIF en Association reconnue d'utilité publique (ARUP) et son changement de nom, 11 voix contre et 2 nuls pour un total de 144 votants sur 280 inscrits.

1477 – * Déclaration à la préfecture de police. *Ancien titre :* SOCIETE DES PERSONNELS ENSEIGNANTS ET CHERCHEURS EN INFORMATIQUE DE FRANCE. *Nouveau titre :* **SOCIETE INFORMATIQUE DE FRANCE (SIF).** *Nouvel objet :* favoriser le développement de l'enseignement et de la recherche ; développer les échanges entre l'enseignement de l'informatique les organismes et laboratoires de recherche et le monde socio-économique, favoriser, par la formation tout au long de la vie et la recherche, l'évolution professionnelles des acteurs publics ou privés, du secteur, oeuvrer à ce que l'informatique et les sciences du numérique contribuent au développement économique et social. *Siège social :* 11, rue Pierre et Marie Curie, 75005 Paris. *Date de la déclaration :* 2 octobre 2012.

Extrait du Journal Officiel

J'aimerais profiter de cette chance qui m'est offerte pour vous raconter les raisons de ce changement et comment nous y sommes arrivés. Mais avant cela, j'aimerais indiquer qu'il ne s'agit nullement d'un travail d'historien et que j'ai bien entendu de nombreux biais. J'ai essayé de ne pas me reposer exclusivement sur mes souvenirs : j'ai pu reconstruire à la fois les événements et les réflexions de l'époque en me basant

1. Professeur des universités, Chaire Unesco RELIA.
2. http://specif.org.

DOI:10.48556/SIF.1024.20.9

sur des comptes rendus, des transparents, des éditos, et les souvenirs des complices de l'époque. Merci en particulier à Jean-Christophe Janodet, dernier secrétaire de SPECIF et premier de la SIF, qui a conservé les nombreux documents permettant de retracer les événements. Les erreurs, omissions, exagérations et interprétations sont bien entendu exclusivement de ma responsabilité. À ma décharge, comme pour tous ceux qui ont vécu cette époque, les sentiments restent assez forts. C'est une vraie aventure collective et une vraie révolution collective qui a eu lieu.

Dans un premier temps, je vous propose une présentation chronologique, en essayant de reconstruire comment nous sommes parvenus à créer la Société informatique de France à travers cinq dates clés. Dans un second temps, j'essaierai d'analyser plus en détail certaines des questions clés qui se posaient il y a 10 ans. Et nous arriverons — ou pas — à la conclusion que ces questions ne sont plus d'actualité. Je finirai en analysant ce qui me parait avoir été le fer de lance de notre aventure.

Grenoble, en février 2011

Une nouvelle équipe se forme pour prendre la direction de SPECIF. Je lis :

— Colin de la Higuera, président ;
— Sylvie Despres, vice-président enseignement ;
— Isabelle Simplot-Ryl, vice-présidente recherche ;
— Laure Petrucci, vice-présidente international ;
— Annie Geniet, secrétaire ;
— Bruno Defude, trésorier.

L'association compte alors 350 membres actifs et 50 laboratoires ou départements sont adhérents en tant que personnes morales. Les finances sont saines et nous venons de remporter les élections au CNU. De quoi se plaint-on ?

Ce qui caractérise cette équipe est qu'elle vient de perdre, en un renouvellement de CA, trois anciens présidents de SPECIF. Dans l'édito du bulletin de mars 2012, j'énumère la liste des sortants : les ex-présidents Christine Choppy, Hervé Martin et Pierre Lescanne ont quitté le CA, ainsi qu'Antoine Petit, Michel Riveill, Stéphane Lavirotte ou Daniel Herman.

La nouvelle équipe est jeune mais voit tout de même revenir des personnes d'expérience qui ont connu et donné de leur temps pour SPECIF depuis près de 20 ans : Florence Sèdes et Max Dauchet. Mais cette nouveauté signifie que l'équipe peut — et doit — travailler en rupture avec le passé. Très vite, le besoin le plus pressant est identifié : « passer ARUP ». Autrement dit, devenir association reconnue d'utilité publique. Derrière cette ambition en 4 lettres on va retrouver trois grandes raisons :

— à l'âge d'or de SPECIF, il y avait 700 membres. En 2011, nous avons le complexe de ne pas être assez représentatifs. Et nous vivons cela — étrangement, avec le recul — comme un véritable frein. Il paraît alors très difficile de

prendre la parole sur la place publique si nous ne représentons pas numériquement la communauté. Et l'analyse de l'époque était qu'en devenant ARUP les deux-tiers de la cotisation seraient déductibles des impôts entraînant mécaniquement une augmentation de la cotisation ou du nombre d'adhérents ;

— la seconde raison de devenir ARUP était que nos partenaires directs (la SMF et la SMAI pour les mathématiques, la SFP pour la physique) l'étaient aussi, et si on veut que l'informatique soit considérée comme équivalente à ces disciplines, il faut absolument passer ARUP ;

— la troisième raison est que l'ASTI vient de prononcer sa dissolution et nous transmet, au moins implicitement, le témoin.

Une parenthèse sur la dissolution de l'ASTI. L'Association des sciences et technologies de l'information (ASTI) avait succédé en 1998 à l'AFCET. Sa dissolution en 2010 s'est accompagnée d'un legs de 10 000€ à la SIF pour financer le travail mené par l'Enseignement public et informatique (EPI) pour l'enseignement de l'informatique dans le secondaire. Si SPECIF avait (peut-être) eu un rôle dans la création ou la recréation de l'option informatique au lycée au XX^{e} siècle, elle n'en a eu aucun dans sa disparition : le sujet avait certes été discuté, mais aucune action tangible pour protester contre cette disparition n'avait eu lieu. Et pendant les années de traversée du désert de l'informatique du secondaire, l'EPI s'est retrouvée bien seule. En fait, elle avait été rejointe par Maurice Nivat — qui allait devenir le président d'honneur de notre conseil scientifique — et s'était illustrée durant la campagne électorale qui a conduit à la présidence de Nicolas Sarkozy en écrivant aux différents candidats, puis en obtenant ainsi un rendez-vous avec le nouveau cabinet. Des méthodes dont nous allions nous inspirer. Nous allions inclure l'EPI dans SPECIF et bénéficier de l'expérience de Jean-Pierre Archambault qui allait nous rejoindre au sein du CA.

L'analyse qui va avoir lieu en 2011, au sein du CA mais aussi des rendez-vous divers de SPECIF pour parler éducation ou recherche, est qu'en fait, la vraie raison de vouloir devenir ARUP est ailleurs : c'est que nous découvrons que les vrais enjeux pour notre discipline ne sont pas au sein de nos laboratoires, de nos départements, de nos instances de gestion de la discipline. En particulier, la question de l'enseignement de l'informatique ne va pas se résoudre dans notre environnement universitaire relativement protégé mais sur la place publique. Or, SPECIF travaillait en interne depuis très longtemps : notre communication était purement entre nous ; c'est envers nos adhérents, exclusivement que nous avions choisi de nous tourner. Malgré des présidentes ou des présidents et des équipes remarquables, le simple fait de rester « entre nous » rendait très peu impactantes nos réflexions et actions, du moins hors, justement, de nos cénacles. Nous souffrions, en quelque sorte, du syndrome du front populaire de Judée.

Dans un tout premier temps, nous pensions qu'il suffisait de s'ajouter la volonté de devenir ARUP, puis de créer en 2011 un conseil scientifique pour pouvoir avancer. Ce conseil représente une étape importante et nous prenons le temps de bien définir

les missions de ce conseil, d'en mesurer l'impact, de lui donner des règles. Mais cela va être en réalité bien plus compliqué.

Paris, décembre 2011



Cette date est une date fatidique. Lorsque je lis mes notes de l'époque, j'y trouve les transparents utilisés pour la réunion du conseil d'administration tenue le 1er décembre. Et on y voit que le choix est fait de créer une société savante et de lui donner un nom. Les noms envisagés plus ou moins sérieusement sont : Société informatique de France (ou variante : de France, française), Société pour Promouvoir, Enseigner, Chercher en Informatique en France (ou variante : l'Informatique), Société pour la Promotion de l'Enseignement, de la reCherche en Informatique en France, Société pour la Promotion, l'Enseignement et la reCherche en Informatique en France, Société pour la Promotion, l'Enseignement et la reCherche en InFormatique, Société pour la Promotion, l'Enseignement et la reCherche en Informatique de France Société des Personnels Enseignants et Chercheurs en Informatique de France, ou encore Société Nationale Informatique Française (SNIF), Promotion de l'Informatique en France (PIF), Promotion de l'Informatique, Télécommunications et Réseaux dans l'Enseignement (PITRE), Fédération Informatique Française de Recherche et Enseignement (FIFRE), Société Informatique de Recherche et Enseignement (SIRE), Professionnels de l'Informatique et Personnels Enseignants (PIPE), Réseau Informatique des Chercheurs et Enseignants (RICE), Chercheurs et Enseignants en Informatique (CEI), Informatique : Recherche et Enseignement (IRE).

En regardant de près ces propositions, on note que nous avions encore l'espoir de garder SPECIF, en changeant uniquement le sens des initiales. Mon souvenir est que c'est à l'issue de cette réunion que la solution est apparue : transformer SPECIF en SIF était beaucoup trop compliqué. Il fallait dissoudre SPECIF.

Toulouse, février 2012

Le congrès de l'association a eu lieu à Toulouse. L'assemblée générale s'annonçait délicate. Certes, il avait été annoncé dans un édito récent que nous proposerions quelque chose de nouveau lors du congrès de Toulouse, mais l'étendue de notre proposition et de sa signification n'était pas encore claire. J'ai un souvenir d'être conscient que quelque chose d'important se jouait et qu'il y avait là un Rubicon à franchir. Cela dit, la proposition a été adoptée en séance, mais le vote n'était pas conforme aux statuts (le vote impliquait quand même la dissolution d'une association vieille de 25 ans, qui se portait très bien, qui détenait à la fois la présidence du CNU et celle de la section 7 du CNRS !). Le conseil d'administration est renouvelé et le bureau qui va porter la création de la SIF est :

Président : Colin de la Higuera
Vice-président enseignement : Philippe Marquet
Vice-président recherche : Jean-Marc Petit
Vice-présidente international : Laure Pétrucci
Trésorier : Bruno Defude
Secrétaire : Jean-Christophe Janodet

On y retrouve Bruno Defude, dont la tranquillité était égale à sa compréhension des enjeux financiers. Mais la plupart des autres postes vont évoluer très vite. Jean-Marc Petit prend la direction des sujets de recherche et assurera en particulier la mise en route du conseil scientifique.

Paris, mars 2012

Une réunion a eu lieu en mars 2012, à Paris, avec les représentants du CA d'une part, « nos » élus d'autre part. Ces dernières années, nous avions eu du succès à la fois au CNU et aux comités du CNRS. Et ma foi, la réunion a été houleuse. C'est dans des situations comme celle-là que la sagesse de Max Dauchet, qui était un des stratèges de la SIF, s'est avérée décisive. Le débat de fond était qu'avec la position du CA nous abandonnions en quelque sorte les élus des listes SPECIF à leur sort. À un moment où, justement, la communauté votait pour ces listes. Certes, nous proposions qu'une fois la SIF créée, une nouvelle association de type « société des personnels » voie le jour mais expliquions que ces deux associations devraient être étanches. Nous étions persuadés qu'il était impossible de maintenir un lien entre les listes et la société savante en création, sous peine de la perdre immédiatement.

Paris, mai 2012

C'était le dernier obstacle. Nous avons préparé avec soin l'assemblée générale extraordinaire de Paris, du 31 mai 2012, à laquelle nous avons voté la dissolution de SPECIF, la création de la SIF et ses nouveaux statuts. Une inquiétude concernait

le quorum. Il nous fallait non seulement avoir assez de voix, mais aussi assez de votants. Je me souviens de l'impression de vivre un moment important et, comme il se doit, la salle a applaudi à la proclamation des résultats.

Analysons maintenant les analyses qui étaient les nôtres durant ce processus. On peut regrouper celles-ci dans trois thèmes : les besoins, l'héritage et la légitimité.

Quels étaient les besoins en 2011 ?

SPECIF avait fêté ses 25 ans un an plus tôt mais ne se portait pas si bien que cela. Pour s'en persuader, on peut revisiter le site de SPECIF qui n'a pas disparu. Les nouvelles de l'époque sont des nécrologies, des résultats électoraux, l'annonce du congrès. Bien entendu, l'activité était plus importante que ça, mais ce n'était pas simple de matérialiser celle-ci. Le président sortant, Pierre Lescanne, avait bien pris conscience du problème mais les conditions n'étaient pas réunies pour mettre en place un autre modèle. Les besoins reposaient sur : l'impasse ressentie depuis longtemps par le CA du rôle de SPECIF, la disparition récente de l'ASTI, la question de l'éducation au lycée, l'impression de parler uniquement entre nous.

Pour faire avancer cette analyse, nous avons alors profité d'un groupe de travail mené par Maurice Nivat et Gilles Dowek en liaison et au nom de l'Académie des sciences : une dizaine à nous rencontrer chaque mois, dans les locaux d'INRIA, pour discuter de la question de l'enseignement de la discipline, et cela depuis le plus jeune âge. Maurice Nivat nous convainquait que savoir programmer était aussi important et élémentaire que savoir lire, écrire ou compter et nous expliquait des algorithmes de comptage de billes découverts par les enfants en école maternelle. Nous avons rencontré largement les sociétés savantes amies (SMF, SMAI, SFP, SEE...). Nous avons discuté avec les sociétés professionnelles avec lesquelles nous allions tisser des liens. Nous avons aussi étudié les modèles des sociétés savantes étrangères (par exemple, nos amis australiens avaient pris en charge l'étude des dossiers des migrants de l'informatique, leur assurant un modèle financier très particulier).

La dissolution de l'ASTI venait aussi nous donner plusieurs points de réflexion. L'ASTI avait voulu créer un rassemblement large. En se définissant comme la société savante des technologies de l'information et de la communication, elle avait eu de mal à mettre en avant la « science ». Le modèle de l'ASTi avait été fédéral. Ainsi, SPECIF était membre de l'ASTI, comme l'EPI et des associations plus thématiques. Dans ce modèle, l'essoufflement était arrivé très vite. L'EPI rentrait dans SPECIF, avec son legs. Quand on interrogeait nos amis de l'EPI sur le nombre de membres, la réponse était plusieurs milliers de lecteurs de la lettre d'EPI (ou sympathisants). Cette réponse avait le don de nous énerver, tant la question du nombre d'adhérents cotisants était importante au sein des conseils d'administrations. Mais là aussi, ce sont des leçons qui allaient nous servir. Pour être complets, il faut mentionner deux autres tentatives de sociétés savantes dont l'analyse alimentait nos réflexions. L'informatique s'était dotée d'une société savante en 1968 (Association française pour

la cybernétique économique et technique ou AFCET) qui disparaît en 1998. LA SEE représente la France à l'IFIP depuis 1998, et devient Société de l'électricité, de l'électronique et des TIC.

L'héritage

Au-delà du besoin de devenir une société savante, il y avait une question majeure à régler, pour y arriver. Celle de l'héritage. Il a vite été clair que la création d'une société savante à partir de rien allait échouer. Il fallait donc la bâtir en consolidant quelques acquis. Des acquis financiers tout d'abord. Lancer une société savante, ça va coûter de l'argent. Nous n'allions pas avoir immédiatement des adhésions, des sponsors, des donateurs. Et SPECIF avait de l'argent... issu en grande partie d'une mission d'observation portée en son temps par Max Dauchet, lorsqu'il était président de SPECIF. Ensuite, des acquis moraux : les adhérents, surtout. Certes, nous plafonnions à 300 adhérents, mais il y avait aussi les personnes morales comme les laboratoires et départements d'informatique un peu partout. Là aussi, il importait qu'ils nous suivent. À cette époque, le terme « numérique » était très volatile et aurait pu sans difficulté être revendiqué par à peu près tout le monde. Ce besoin d'héritage nous poussait vers l'unique solution de transformer SPECIF en SIF.

La légitimité

La légitimité avait toujours été un enjeu pour SPECIF. Et l'importance croissait avec une ambition bien plus importante pour la SIF. Nous étions conscients que les sociétés savantes en face de nous s'appuyaient sur une réputation, sur un historique long. Avions-nous le droit de nous autoproclamer société savante de l'informatique ? D'écrire, de dire « Nous, la société savante de l'informatique pensons que... » ?

Ici encore, on pouvait penser qu'en changeant simplement les statuts de SPECIF et en nous appuyant sur nos adhérents, sur nos laboratoires, nous allions y parvenir. Mais notre analyse était que cela ne suffirait pas, pour au moins deux raisons. La première était que SPECIF n'avait pas participé aux débats publics importants de société concernant l'informatique. Certes, nous en avions discuté entre nous mais bien plus rarement nous avions cherché à intervenir dans le débat public. La seconde était « la question des listes ». Depuis plusieurs mandatures (1995), SPECIF présentait ses listes au CNU en 27^{e} section. Elle gagnait régulièrement chez les professeurs et obtenait chez les maîtres de conférences des sièges qui lui permettaient — souvent — de présider cette institution. Avec l'appui des membres nommés, cela faisait de SPECIF un rôle important dans la gestion des recrutements et des carrières. Bien entendu, si cela correspondait à un besoin exprimé par des collègues, c'était aussi un sujet clivant dans les discussions avec les collègues syndiqués. L'analyse du conseil d'administration était que si la SIF se voulait la société savante de tous les informaticiens, elle ne pouvait pas soutenir — et encore moins organiser — une liste

particulière. En 2012, SPECIF présentait également des listes de candidats aux élections du Co-CNRS ; là aussi, avec succès. La légitimité dépendait également de la relation à définir avec les deux acteurs de référence qu'étaient l'INS2I et INRIA. Il fallait absolument les convaincre du bien-fondé de la démarche, de la nécessité de statuts qui rendent la SIF réellement indépendante de ces « tutelles » tout en obtenant leur soutien qui était vraiment indispensable : soutien financier à travers l'adhésion des laboratoires comme personnes morales, soutien politique en conviant la SIF dans les grandes alliances sur le numérique qui se dessinaient alors. Et pour des raisons que l'on comprend, l'abandon du rôle de co-gestionnaire de la discipline n'allait pas être simple à vendre. Là aussi, les discussions se sont toujours bien passées et les soutiens apportés par nos interlocuteurs au CNRS-INS2I et à INRIA ont été sans conteste un atout majeur pour la suite.



La légitimité s'exprime aussi à travers le choix des mots. Les mots « à la mode » en 2012 étaient numérique, code et *coding*. Que nous avions envie de remplacer par informatique et programmation ; avec de nombreux arguments que nous avions l'habitude de partager... entre nous !

Mais le positionnement avait changé : si nous voulions être écoutés, participer à des débats de la société civile (comme l'alliance EDUCNUM mise en place par la CNIL et qui s'est avéré décisive pour trouver des alliés), il fallait choisir entre discuter des mots-sujets qui intéressaient la société (le numérique, le code, le *coding*) mais en imposant notre vocabulaire, faire de même en expliquant notre ontologie, ou descendre dans l'arène en acceptant les règles du jeu, c'est-à-dire en utilisant les mots que les gens connaissaient, au risque de leur apporter de la légitimité. Ce débat a eu lieu et aura toujours lieu. La première solution consiste à ne parler qu'à des gens qui veulent nous écouter, et très vite, à parler entre nous. La troisième est une position difficile à tenir pour une société savante. La deuxième est celle que nous avons choisie mais ne va pas de soi ; elle consiste à partir d'un positionnement à la fois convainquant pour nous, mais aussi pour les autres. Le conseil scientifique a bien entendu été mis à contribution pour poser les définitions puis publier dans 1024 et binaire les articles de fond si nécessaire.



En lisant cet article, on peut penser que la démarche était analytique, que nous avions toujours un temps d'avance. Ce n'était pas du tout le cas : descendre dans l'arène publique, c'est aussi en accepter les règles. Une information sur l'enseignement de l'informatique ? Il faut réagir très vite si on veut que cette réaction soit audible. Un débat public et une salle qui réagit négativement à une position *made in* SIF ? Il faut choisir une autre approche pour faire passer le message, utiliser d'autres arguments et on ne peut pas toujours attendre la prochaine réunion du conseil scientifique ou du conseil d'administration. Une interview dans laquelle il faut imposer un sujet polémique pour être certain qu'elle sera publiée : oui. Il y a alors une prise de risque ; celle d'aller plus vite que son bureau, que son CA. Raison de plus pour communiquer abondamment.

Quelles ont été les « retombées » de cette approche ?



En premier lieu, nous avons trouvé des alliés. À l'époque, dans des réunions parfois complexes, nous étions très minoritaires à penser qu'apprendre l'informatique était utile et nécessaire à tous. Et comme en plus l'immense majorité de nos interlocuteurs n'avaient jamais programmé de leur vie, il nous a fallu trouver de nouveaux arguments pour convaincre.

En second lieu, nous avons appris à parler en public, au public. J'ai le souvenir d'avoir très vite abandonné des arguments qui fonctionnaient très bien entre nous, mais qui, quand ils étaient énoncés devant une salle pas du tout acquise à la cause de l'éducation à l'informatique, suscitaient des incompréhensions, des réactions hostiles difficilement reprochables ; d'autres arguments faisaient mouche.

En troisième lieu nous avons compris que ce qui déterminait le plus cette capacité de communiquer était la médiation scientifique. Effectivement, c'est plus facile de convaincre quelqu'un au ministère quand on a passé un samedi après-midi à expliquer l'informatique à la sortie d'un hypermarché !

En quatrième lieu, nous avons rendu la SIF visible. En choisissant délibérément d'axer notre communication sur tout le monde plutôt que sur nos adhérents, nous prenions un risque envers ces adhérents. Mais nous nous obligions à communiquer autrement. À l'époque, le rêve était de voir 1024 circuler entre les collègues professeurs de lycée ! Mais je pense sincèrement que nos adhérents nous ont suivi. Lorsque nos tribunes étaient publiques, ils pouvaient se reconnaitre en elles et savoir que leur voix était portée sur la place publique.

En cinquième lieu nous avons participé aux débats de l'époque. Il est bien entendu impossible de démontrer que ces prises de position se sont avérées décisives, mais j'ai quand même le souvenir que nos tribunes étaient lues, discutées, et que nos arguments pouvaient revenir au détour d'une visite à un ministère.

Si nous fêtons aujourd'hui les 10 ans de la SIF, c'est que, probablement, quelques-unes des idées portées dès 2012 étaient bonnes. Si je dois aujourd'hui mettre en avant quelques recettes, en voici trois.

L'importance de la médiation scientifique. En mettant l'emphase sur ce sujet, nous annoncions clairement notre intention d'aller vers le grand public. Nous affichions notre volonté de communiquer. Cet axe, porté magistralement par Sylvie Alayrangues, a su bénéficier de deux efforts formidables : 1024 a été porté par Eric Sopena. Depuis l'adaptation des fichiers latex (et il me semble même que les fichiers de départ étaient TeX), jusqu'à l'établissement d'une ligne éditoriale tout à fait nouvelle. Peu après, c'est Serge Abiteboul qui lançait Binaire. Ce blog, hébergé par Le Monde, est vite devenu une référence scientifique pour l'informatique.



L'enseignement de l'informatique comme vecteur de rassemblement. La « grande cause » qu'était l'enseignement de l'informatique, à tous les niveaux, a été un excellent élément de rassemblement. Mais il est important de dire que l'accord n'allait absolument pas de soi. Les points de vue de départ étaient très différents et les sujets de friction étaient nombreux (par exemple, la place du numérique dans l'informatique). Puis, ce qui a permis à ce sujet d'avancer est clairement la grande valeur des

personnes. Philippe Marquet, vice-président, a eu le mérite de porter ce sujet, avec des personnalités très fortes systématiquement liées à ces questions. Les membres du conseil scientifique, en particulier, forts de l'écriture du rapport « L'enseignement de l'informatique en France – Il est urgent de ne plus attendre » ont été systématiquement mis à contribution pour écrire, participer à des délégations, rencontrer, discuter...

Actions politique et de médiation. Il y a aussi une méthode qui se voulait en rupture avec ce qui s'était fait jusque là. Lorsqu'un problème était identifié, deux actions étaient déclenchées en parallèle : politique et médiatique. Typiquement, la première consistait à prendre un rendez-vous dans un des ministères. La deuxième consistait à écrire une tribune dans un journal, ou dans un média en ligne, ou encore réaliser une interview. Je reste persuadé que cette action médiatique (où nous avons dû tout apprendre) était essentielle et j'ai plus tard eu confirmation que nos prises de position publiques avaient un impact.

Le 31 mai 2012, naissait la Société informatique de France. Nous fêtons son dixième anniversaire cette année. Je lui souhaite une adolescence aussi réussie que son enfance.

Table ronde des anciens présidents

Fabien Tarissan [1]

Après avoir expliqué l'histoire de la création de la SIF par Colin de la Higuera, les organisateurs du congrès ont souhaité donner la parole aux trois présidents qui ont œuvré pendant ces 10 dernières années. Cette table ronde est animée par Fabien Tarissan, ancien vice-président médiation de notre association.

SIF (Fabien Tarissan) : « *Colin, tu as évoqué dans ta présentation de l'histoire de la création de la SIF* [2] *ce qui s'était joué en 2012 au moment de la rupture avec SPECIF et de la création de la SIF. Mais cela ne nous dit pas comment cela s'est joué. Face à l'ambition de ce que devait être la SIF, face aux enjeux à ce moment-là, comment la SIF s'est-elle construite sous ta présidence, en termes de gouvernance ou d'animation de cette communauté ? »*

CH (Colin de la Higuera) : J'ai cité tout à l'heure un premier ministre anglais célèbre ; je citerais une deuxième personne, c'est Margaret Thatcher qui disait « *We have a democracy and I run it* ». C'était une plaisanterie pour expliquer qu'en fait, il y a des débats, mais si les besoins sont clairs pour tout le monde et que tout le monde voit qu'il faut absolument qu'on obtienne quelque chose, on arrive à trouver des positions de consensus, et elles sont solides. Quand tu sais que la semaine après, on rencontre quelqu'un ou un cabinet, quand tu sais qu'il y a une interview, la position, il faut la trouver, donc tu vas réussir à obtenir cette convergence sur les instruments... Je crois que ça a encore été grâce à Max Dauchet ça : à la fin de SPECIF, on avait mis en place le conseil scientifique, avec des textes extrêmement serrés, combien de personnes, combien de places pour chaque catégorie, qui, quoi,

1. Enseignant-chercheur, ENS Paris-Saclay.
2. https://doi.org/10.48556/SIF.1024.20.9.

DOI:10.48556/SIF.1024.20.21

comment, enfin toute une série de choses. Nous avions passé du temps à préparer ce texte, et ce conseil scientifique, une fois qu'il a existé, il fallait lui donner des choses à faire, donc des sujets de réflexion, avec un CA qui était peut-être un peu plus dans l'action et peut-être un peu moins dans la réflexion, et le conseil scientifique qui était plus dans la réflexion pour trouver et enrichir des positions qui, ensuite, pouvaient être tactiquement défendues. Les rôles, comme ça, se sont retrouvés... Bien sûr, institutionnellement, il faut avoir les bons objets pour pouvoir avancer.

SIF : « *Sur ces premières actions de la SIF que tu évoques (ces rencontres, ces interviews...), comment avez-vous avancé ? Max a pu dire qu'à cette époque vous vous sentiez comme des gamins. Est-ce que le passage à SIF a fait que, très rapidement, vous ne vous êtes plus sentis comme des gamins ? »*

CH : Non, tu allais au ministère, et tu étais là, en train de te dire : « En fait, on n'est pas du tout sûrs... ». Quand il fallait avoir des points de vue sur l'enseignement, sur l'ordre des choses entre le CAPES et l'agrégation, toute une série de sujets, tu y allais, on en avait discuté, mais t'étais pas sûr... On n'était pas sûrs de nous, mais, en fait, on savait aussi qu'il n'y avait personne d'autre autour qui allait être plus sûr que nous. Donc la partie démocratique était importante, c'est-à-dire qu'il y avait eu suffisamment de conversations pour que, globalement, on propose une position commune, et ça se passait bien.

SIF : « *Et comment a été accueillie l'arrivée de la SIF par nos partenaires naturels (CNRS et INRIA notamment mais aussi les autres sociétés savantes, etc.) ? Est-ce que cela a été facile de poser l'existence de la SIF ? »*

CH : On a juste passé énormément de temps à rencontrer et à préparer à l'avance, donc les institutions, on allait les rencontrer. Pour le CNRS et INRIA, si possible les deux en même temps, ce qui est toujours un exploit, et on arrivait effectivement à une position avec elles ; idem pour à peu près tous les autres groupes, les sociétés savantes... Le plus difficile, ça a été avec les élus de SPECIF mais ça se comprend très bien. On a beaucoup discuté. Pour les adhérents, sur le terrain, c'était la sensation qu'on avait quand on était dans les réunions habituelles de SPECIF, qui correspondent à celles d'aujourd'hui, donc il y avait une salle, tu discutais avec les gens, et t'avais quand même cette sensation que les gens étaient avec nous ; probablement pas tous. Il y avait des gens pour qui le rôle le plus important, c'était autre chose, et nos histoires de médiation, toutes ces questions-là qui nous intéressaient, ce n'était pas SPECIF.

SIF : « *Sur un plan plus personnel, tu as été le premier président de la SIF, on est maintenant dix ans plus tard, quel regard portes-tu maintenant sur cette association, sur ce qu'elle a pu accomplir ? Est-ce qu'il reste des choses qui étaient dans les ambitions de ces premières années sur lesquelles il faut encore travailler ? Comment tu vois le devenir de la SIF sur ce qu'il lui reste à faire ? »*

CH : D'abord, bravo pour tout ce qui a été fait, je suis très impressionné. Moi, j'aimerais bien un film sur Netflix, c'est là qu'il faut mettre la barre, c'est-à-dire que si on veut convaincre les gens, c'est à ce niveau-là, qu'il y ait des séries à la télévision qui racontent la vie des informaticiens. C'est là qu'il faut aller gagner la bagarre. À un moment donné, on a essayé de travailler avec les médias, les chaînes de télévision, mais on n'a pas réussi. Mais l'ambition doit être, là-dessus, très haute. Sur l'enseignement de l'informatique, le CAPES et l'agrégation, c'est très bien pour former les futurs informaticiens, mais ce qui me préoccupe, c'est que l'informatique est globalement absente dans les INSPE pour tous les autres enseignants —sauf exception !—, et que la culture informatique de l'enseignement de base en France n'a pas bougé en dix ans. Ce n'est pas de la faute de la SIF : on connaît les problèmes, mais là, il y a un chantier absolument gigantesque devant nous. Donc oui, il y a des choses à faire, vous allez vous amuser.

SIF : « *Merci Colin. Avançons maintenant un peu dans le temps et tournons-nous vers Jean-Marc Petit. Jean-Marc, tu prends la présidence du CA en 2015, c'est à dire trois ans après la naissance de la SIF, à un moment où il y a quand même un certain nombre de choses qui ont été construites et mises en place. Une partie de l'action que tu as entreprise, c'était aussi de professionnaliser tout le travail de la SIF et de poursuivre ce travail d'unité entamé précédemment. Est-ce que tu peux nous raconter un peu ce qu'il s'est joué quand tu étais président, quelles ont été tes motivations, quels ont été les dossiers et les moments marquants de ta présidence ?* »

JMP (Jean-Marc Petit) : C'est avec beaucoup de plaisir que je vais tenter de répondre à tes questions. De façon très personnelle et un peu anecdotique, je me rappelle ce qui a déclenché mon engagement. Au congrès de SPECIF à Grenoble en 2011, Gilles Dowek avait fait un exposé sur la place de l'informatique dans les sciences [3]. Cet exposé sur l'épistémologie des sciences m'avait marqué : ce qu'il y avait dit apparaissait comme une évidence, et donnait une forme de fil directeur pour l'unité de la discipline. C'est, je crois, la boussole qui m'a guidé durant toutes les années qui ont suivi au sein de la SIF.

De façon plus précise maintenant, le premier défi était d'expliquer aux informaticiens et informaticiennes, pas aux autres collègues d'autres disciplines, qu'il existait un plus petit dénominateur commun qui était fort, puissant, sur lequel on pouvait se retrouver. L'exemple typique, qui est encore très vrai pendant ce congrès des 10 ans de la SIF, c'est autour de la médiation scientifique : on peut faire de la théorie des graphes, des systèmes distribués, de l'IHM, et puis trouver des sujets communs pour expliquer la science que l'on fait. Donc le premier défi, c'était d'arriver à expliquer que nous appartenions à une même grande famille.

3. « Sur l'apport de la philosophie des sciences aux sciences elles-mêmes », compte rendu d'Annie Geniet, pages 22-24, https://www.societe-informatique-de-france.fr/bulletins-specif/specif065.pdf.

Le deuxième défi était de rassembler la communauté dans un paysage qui était extrêmement morcelé. Les informaticiens n'ont pas attendu la SIF pour s'organiser en créant d'innombrables associations ou groupes de travail. Dans l'enseignement secondaire, il existait par exemple l'Union des professeurs de classes préparatoires scientifiques (UPS [4]) et l'association Enseignement public & informatique (EPI [5]). Dans l'enseignement supérieur et la recherche, il existait trois types de structuration : les associations thématiques de l'informatique formant un réseau très dense, les GDR du CNRS et les sociétés savantes amies. Le rôle de structuration joué par le conseil des associations de la SIF a été primordial. Enfin dans l'industrie informatique, elle était aussi extrêmement structurée, avec par exemple l'association des grandes entreprises et administrations publiques françaises (CIGREF [6]) ou Pascaline devenue Talents du numérique [7] soutenue par le Syntec numérique (Numeum [8]).

Dans ce paysage, il fallait trouver une façon de se regrouper. Autour de ça, l'ambition a été de dire que « *l'institution passe avant les individus* ». La SIF devait être plus forte que les gens extrêmement brillants que l'on pouvait mettre en avant et qui représentaient la discipline souvent en leur nom. Là, je suis quand même obligé de citer les deux personnes qui, à mes yeux, ont eu un rôle tout à fait majeur pour le positionnement de la SIF, à savoir les deux présidents du conseil scientifique de l'époque, Serge Abiteboul et Gilles Dowek. Ils ont vraiment implémenté cette ambition. Ils ont eu l'intelligence de mettre leur personnalité en retrait par rapport à l'institution et de jouer collectif. Nous avions beaucoup de « fines gâchettes », souvent brillantes et passionnantes. Un des objectifs a vraiment été de les faire « tirer dans la bonne direction », celle qui favorisait l'unité de la discipline.

Dans un paysage extrêmement fragmenté, il a aussi fallu pousser des principes de bon sens : faire des choses simples, ne pas être dans l'incantation, prendre des sujets, proposer des choses, se tromper, ce n'est pas grave, mais faire des propositions et avancer.

SIF : « *On perçoit bien cette nécessité de rassembler autour de sujets communs mais, justement, quels sujets ont été fédérateurs ?* »

JMP : Il y avait par exemple ce sujet important, lancé par Colin, autour de la parité homme-femme. Là encore, il fallait ne pas être dans l'incantation, faire les choses avec sincérité. Quand j'ai pris la présidence de la SIF, je l'ai prise aussi parce que Christine Froidevaux avait accepté d'être présidente-adjointe. Le rôle des femmes dans la genèse de la SIF a été déterminant. On avait peut-être l'habitude d'être un peu entre nous, et elles ont eu un rôle tout à fait crucial en termes d'ouverture, de

4. https://ups-cpge.fr.
5. https://www.epi.asso.fr.
6. https://www.cigref.fr.
7. https://talentsdunumerique.com.
8. https://www.syntec.fr.



posture et d'action. Souvent très discrètes au début, elles montaient en puissance avec le temps. Je me rappelle aussi que Christine pouvait avoir « peur » de telle ou telle personne au début de son mandat, mais qu'à la fin, c'était peut-être bien les autres qui avaient peur d'elle ! Je mentionne Christine, je mentionnerais aussi Catherine Berrut, qui sont deux femmes d'une ténacité et d'une efficacité absolument redoutables pour l'intérêt général.

SIF : « *D'autres personnes que tu voudrais nommément remercier pour leur implication ?* »

JMP : Oui. C'est une période qui a été extrêmement riche, avec un soutien très fort de l'INRIA et du CNRS, en l'occurrence de Michel Cosnard, Antoine Petit et Michel Bidoit. On les rencontrait régulièrement, ils nous transmettaient un optimisme — chacun à leur façon — tout à fait significatif. La SIF s'est faite aussi grâce à eux. Dix ans après, les personnes ont changé mais je constate que la SIF est une zone de discussion entre toutes ces personnes, le CNRS, l'INRIA ou les universités diverses et variées, et il me semble que dans le paysage actuel, c'est tout à fait important de le préserver et de l'amplifier.

Enfin, je souhaitais citer quelques noms avec qui j'ai été plus proche et les remercier : Bruno Defude, trésorier intemporel de SPECIF puis de la SIF ; Pierre Gançarski, notre ineffable Pierre, jamais avare de joutes verbales pétillantes ; Jérôme Durand-Lose, la force tranquille ; Michel Raynal, le sage parmi les sages et enfin l'hyperactive Florence Sèdes. Je m'excuse auprès de toutes celles et ceux que je n'ai pas cité, souvent membres du CA de la SIF, qui ont contribué dans la bonne humeur à faire grandir la SIF.

SIF : « *On voit à travers ton témoignage que l'histoire de la SIF n'est pas seulement celle d'une association mais aussi et surtout une histoire de celles et ceux qui la font vivre !* »

JMP : C'est une belle histoire même si l'avenir, à mon sens, reste incertain : on pourrait se demander pourquoi on ne fête pas les trente, quarante ou cinquante ans de la SIF. C'est une bonne question me semble-t-il que de regarder dans le rétroviseur et de bien comprendre pourquoi.

Bon, on fête les dix ans, c'est déjà pas mal ! Il faut préserver ce bien commun et se méfier des vieux démons qui pourraient éventuellement ressurgir à droite ou à gauche.

SIF : « *Tu as mentionné la question de l'industrie, du rapport de la SIF avec le monde des entreprises... C'est un dossier sur lequel la SIF semble moins visible et les liens avec le monde économique ne sont pas forcément encore très mûrs. Comment, d'après toi, pourrions-nous tisser des liens avec le monde des entreprises qui ont aussi besoin, sûrement, d'une société savante qui les accompagne ?* »

JMP : Il me semble que la feuille n'est pas blanche mais il y a une telle activité autour du numérique, aujourd'hui, dans le monde industriel que c'est difficile pour la SIF de se placer. Il y a quand même des contacts avec les grandes associations industrielles mais elles n'ont pas vraiment besoin de nous... Alors, quand elles peuvent avoir besoin de nous, il faut être là, mais c'est beaucoup d'énergie pour arriver à les attirer. Je vois que dans le conseil scientifique, le collège des industriels est plus difficile à constituer que les autres.

Pour un industriel, participer au CS ou au CA, c'est s'engager à passer du temps pour discuter parfois de sujets qui peuvent un peu traîner en longueur... Il y a une notion d'efficacité dans notre rapport au temps que nous perdons parfois. Il faut maintenir des liens autant que possible mais notre bataille n'est au fond pas avec les industriels. Notre bataille est d'exister, de faire que l'informatique existe au même titre que les mathématiques, la physique, la chimie, et ça, tout le monde y a intérêt.

SIF : « *Je trouve que tu as soulevé un point très important. Qu'est-ce qu'une structure comme la SIF peut apporter à ce monde économique ?* »

JMP : La légitimité, le fait d'être capable d'amener des spécialistes sur les questions qui tourmentent la société. Quand on voit le niveau de discussion sur l'IA aujourd'hui, notamment dans les médias, c'est quand même assez catastrophique. La SIF doit être là pour ramener de la rationalité et porter la voie de la science.

SIF : « *Merci Jean-Marc. Terminons ce tour d'horizon des anciens présidents de la SIF avec Pierre Paradinas. Pierre, quand tu arrives à la présidence de la SIF en 2018, tu reprends une structure qui est un espace de réflexion collective et d'action. Ce qui me semble un peu spécifique à ce moment, c'est que les dossiers sur lesquels*

travaille la SIF se sont complètement diversifiés. On a beaucoup parlé de l'enseignement, qui semble être une pierre angulaire au moment de la création de la SIF, mais quand tu prends la présidence, il n'y a pas que l'enseignement, il y a beaucoup de dossiers, par exemple le dossier ARUP, ouvert sous Colin, est passé par Jean-Marc et termine avec toi. Qu'est-ce que tu peux nous dire de cette animation de la communauté au moment où la SIF se diversifie ? »

PP (Pierre Paradinas) : Je ne sais pas ce qu'ont fait mes prédécesseurs sur l'ARUP mais moi, je n'ai rien fait, parce qu'effectivement, c'est Élisabeth Murisasco qui a piloté le dossier de demande... moi, j'ai rien fait, j'ai simplement eu un mail, un jour, qui disait « *C'est paru au Bulletin officiel, ça y est, on est ARUP ; génial, on va payer moins d'impôts, les caisses de la SIF vont se remplir* ». Tout ça pour dire qu'effectivement, j'ai rien fait, et sur pas mal de dossiers, j'ai pas fait grand-chose parce que c'est sympa de présider la SIF... Parce qu'en fait, on préside, on ne dirige pas vraiment : c'est une vraie équipe avec les personnes du CA et du CS, de temps en temps avec des personnes de l'extérieur — pas assez à mon goût — qui sont là aussi, qui travaillent pour la communauté, et donc c'est extrêmement sympa, extrêmement agréable, de présider la SIF. N'hésitez pas à être candidat quand Yves Bertrand aura terminé son mandat et n'hésitez pas aussi à rejoindre le CA de la SIF. Tout cela, c'est un peu vrai et un peu faux parce que Jean-Marc m'avait vendu *la présidence* en disant : « *Tu verras, il y a pas grand-chose à faire* », et manque de bol, au moment où je suis arrivé, c'était le moment où il y a eu la création du CAPES, et il y a un truc qui a changé, je pense, dans la façon dont la SIF a travaillé, et on en a récolté les fruits — désolé, je vais rester sur l'enseignement — de ce qu'avait semé tout le monde avant, c'est-à-dire que cette fois-ci, on allait plus rencontrer les cabinets, on allait plus voir les directions, on allait voir les services, et on allait voir les services avec des armes parce qu'en fait, tout le monde avait bossé avant, et quand on rencontrait les services, on leur expliquait ce qu'on voulait, comment on le voulait, et on leur disait que s'ils avaient besoin de monter un diplôme inter-universitaire, sur toute la France, pour former tous les profs qui allaient arriver pour enseigner NSI, en fin de compte, ils avaient besoin de nous, donc on les a aidés. En échange, on a essayé de négocier certaines choses, on n'a pas réussi à tout négocier, mais je pense que le discours a changé, c'est-à-dire que Max Dauchet, qui était un gamin quand il allait voir les ministères... Bon, moi, j'étais un peu plus vieux que lui, mais c'était quand même relativement facile et agréable de rencontrer les services et de parler... Alors, on ne parle pas à égalité avec les services mais on avait au moins des arguments sérieux, hyper bien préparés par toutes les équipes qui travaillaient soit dans le CA, soit dans le CS, soit autour, et ça, je pense que ça a été aussi un changement, parce que je me souviens que mes prédécesseurs allaient voir les cabinets, et nous, on n'allait pas voir les cabinets, on allait voir les directions. Donc ça, c'était sur le côté enseignement.

Sur les autres dossiers, on leur a parlé d'une maison de l'informatique et du numérique en France [9].

SIF : « *Justement, peux-tu nous parler de cette question du patrimoine ? Car tu as porté plusieurs projets autour de cette question de la transmission, qui t'est chère.* »

PP : C'est vrai qu'il y a un point dont on n'a pas discuté hier, quand on a parlé de la médiation. On est tous persuadés que la médiation est hyper-importante mais on n'a pas parlé des problèmes de patrimoine, et même si on est encore tous très jeunes, même si la science est relativement jeune, c'est une science qui va extrêmement vite, il y a plein de choses qui disparaissent, il y a même un certain nombre de collègues qui ont déjà disparus, et, en fait, on ne s'intéresse nulle part au patrimoine.

Il y a eu un excellent numéro de la CILAC [10], qui est une revue sur le patrimoine industriel, qui parle de cette question du patrimoine informatique et numérique. En fait, on est assez délinquants, aujourd'hui, en France — le Musée des arts et métiers, le conservatoire, en premier — parce qu'on ne s'occupe pas complètement de ce patrimoine, on pense que ce qui est important, c'est de collecter des machines. Alors, oui, c'est important de collecter des machines, mais des machines dans un couloir, ce n'est pas très intéressant, et donc là, il y a une délinquance sur ces manques. La SIF essaye d'alerter, mais là, on est des gamins parce qu'on n'est pas entendus par le Ministère de la culture, on est pas entendus par le Ministère de l'enseignement supérieur sur ces questions-là. Donc là, il y a encore un grand chemin à faire, de la même façon que, vis-à-vis des entreprises, il y a un grand chemin à faire. Je pense qu'il est du ressort, bien évidemment, de la SIF, mais il est aussi du ressort des industriels, qui sont assez grands pour savoir ce qu'ils ont à faire quand ils ont à le faire, pour savoir quand ils doivent être courageux technologiquement. Ils sont plutôt courageux d'un point de vue financier mais le courage technologique, ce n'est pas toujours ce qui les motive.

SIF : « *Merci Pierre. En complément de ces témoignages, il est intéressant de connaître le regard que certains compagnons de route de la SIF portent sur ces années et ces actions. Robert Cabane, toi qui a été inspecteur général et nous a accompagné sur de nombreux dossiers liés à l'enseignement, que peux-tu nous dire sur ces arrivées de la SIF dans le paysage qui était le tien à l'époque et comment cette société savante a évolué ?* »

RC (Robert Cabane) : Je voulais remercier les différents présidents pour ces aperçus d'une grande véracité. J'étais à cette époque inspecteur général de l'éducation nationale et donc, j'ai vu les mêmes événements de l'autre côté de la barrière, et ce qui était extrêmement net, c'est que dans le processus qui a amené, non seulement, à la création, mais aussi à l'édification d'une société savante, l'autorité de la SIF au

9. http://musee-informatique-numerique.fr.
10. https://www.cilac.com/numero-73-patrimoine-industriel-informatique.

sujet de l'informatique a grandi à une vitesse stupéfiante. À l'époque de SPECIF, le Ministère de l'éducation nationale ne voyait aucune autorité particulière sur le domaine et se revêtait d'une autorité propre sur ce qu'il aimait bien, c'est-à-dire le numérique, et par extension, ce qu'il n'aimait pas, c'est-à-dire l'informatique. Tout le basculement, ça a été d'avoir en face un interlocuteur absolument incontestable, une société savante articulée avec la communauté des enseignants-chercheurs ainsi qu'avec l'Académie des sciences ; une société savante qui parle de l'informatique d'une façon parfaitement raisonnable et crédible, et par rapport à laquelle les têtes pensantes du Ministère de l'éducation nationale ne valent pas tripette parce que leur réflexion ne dépasse pas des slogans sur le numérique. Et donc, il y a eu un véritable basculement qui a conduit à la situation actuelle où, en définitive, c'est la SIF qui mène la danse. Cela a pris un certain temps et je crois que rien n'aurait pu se faire sans le passage à l'étape société savante reconnue d'utilité publique, donc j'ai beaucoup appuyé cela, et même pour une raison tout à fait interne : pour un ministère, avoir des interlocuteurs, c'est essentiel, parce que dans une communauté horizontale où tout se vaut, on ne sait pas bien s'il faut parler à telle ou telle personne, parce qu'on risque même de faire des catastrophes, tandis que quand on a un interlocuteur, c'est génial. Donc ça a été un processus indispensable, assez difficile ; c'est clair qu'il y a eu beaucoup de travail pour les présidents et tout leur entourage, et félicitations à tous.

SIF : « *Du coup, vous, les trois ex-présidents du CA, vous avez senti ce passage-là ? Parce que Colin disait tout à l'heure : « On va chercher, on va discuter avec les interlocuteurs », et là, on entend une parole qui dit : « On va aller chercher les gens de la SIF », alors est-ce que vous avez senti ce basculement, cette identification ? »*

PP : Oui. J'ai le 06 d'Édouard Geffray, le directeur de la Direction générale de l'enseignement scolaire (DGESCO), et il répondait à mes mails et mes appels sans difficulté. Ce n'était pas toujours lui en direct, bien évidemment, parce qu'il est un peu occupé, ce monsieur, mais il m'est arrivé d'envoyer des courriers au nom de la SIF à des gens très importants, qui ont des emplois du temps extrêmement chargés et qui ne répondent jamais, ce qui est normal parce qu'ils sont très occupés, et là, la DGESCO répondait. Antoine Petit [11] nous répond aussi, c'est un copain, mais... Il y a plein de gens qui ne répondent pas, alors je pense qu'ils sont beaucoup occupés, ou qu'ils ont des problèmes de mails, je ne pense pas qu'ils soient impolis ou que la SIF ne soit pas importante.

CH : Moi, je découvre ce que dit Robert Cabane, et je trouve cela génial. Je pense que quand je suis parti, nous n'en étions pas encore tout à fait là ; c'est arrivé après, donc c'est très bien.

11. Directeur du CNRS.

JMP : Moi, je dirais qu'il y a eu des frémissements, on a commencé à être sollicités directement, l'institution plus que les individus, mais je pense que c'est à partir de Pierre que ça s'est généralisé.

CH : Moi, j'aimerais bien que dans dix ans, il y ait un Robert Cabane de l'industrie qui dise ce que Robert Cabane a dit aujourd'hui pour l'enseignement. Ce serait important. Ensuite, sur le sujet de l'enseignement, il faut faire attention. On peut avoir l'impression qu'on a gagné parce que le CAPES et l'agrégation sont arrivés, mis il y a vraiment des enjeux au-delà de ça : les enjeux de l'informatique pour tous, tous les enseignants, la présence de l'informatique de manière systématique dans les INSPE, pour qu'elle soit enseignée de façon générale, c'est un combat encore à mener. Celui qui m'intéresse en ce moment, c'est celui de l'intelligence artificielle, et encore plus que celui de son enseignement, parce qu'il y a un ouragan sur l'éducation qui est sur le point d'arriver et si on ne prépare pas les enseignants, il va y avoir un vrai souci. Donc là, les chantiers sont énormes, mais là aussi, cela justifie qu'il faille que demain, dans dix ans, tous les enseignants de France aient eu une petite formation à l'informatique, à ses usages, mais aussi à comment cela fonctionne, pour pouvoir réagir à des outils qui vont arriver dans leur environnement. Ça serait ça, les deux gros défis pour moi.

JMP : Tout a été dit, plus ou moins. Les défis pour consolider l'existant, ça me paraît être la nécessité numéro un. Les efforts que Max a déployés pour créer le conseil scientifique indépendant du conseil d'administration, source de réflexion et de proposition sur des grands sujets, il faut qu'ils perdurent, que ça continue, il faut aller dans cette voie-là. Donc, c'est continuer et transformer l'essai autour du CAPES et de l'agrégation, sait-on jamais, et ne pas tomber dans une forme de routine ankylosante.

SIF : « *Ce sera le mot de conclusion, on peut remercier les trois ex-présidents du CA pour ce témoignage.* »

Une association reconnue d'utilité publique

Élisabeth Murisasco [1]

La reconnaissance d'utilité publique d'une association est une étape importante dans sa vie et représente un agrément de l'État par lequel il montre sa volonté de valoriser et légitimer l'activité de l'association. La SIF a reçu ce label en 2018 après un long parcours administratif et les conséquences sont riches :

— la France est dorénavant dotée d'une société savante en informatique comparable à ce qui existe pour les autres sciences et dans d'autres pays ;

— la science informatique prend toute sa place et elle s'éloigne d'une vision tournée exclusivement vers les usages du numérique ;

— le rôle de la SIF est renforcé et reconnu pour ses actions : (1) éclaircir le monde numérique de notre époque, et apporter des réponses aux nombreuses interrogations que ce nouveau monde soulève, notamment sur des questions d'éthique ou d'enseignement des sciences au XXI^{e} siècle, (2) développer les échanges entre les établissements d'enseignement de l'informatique, les organismes et unités de recherche et le monde socio-économique, (3) favoriser le développement de l'enseignement et de la recherche en informatique, (4) œuvrer pour que l'informatique et les sciences du numérique contribuent au développement économique et social.

La Société informatique de France (SIF) a été fondée en 1985 comme une association loi 1901 sous le nom de SPECIF – Société des personnels enseignants et chercheurs en informatique de France. Le conseil d'administration de SPECIF a voté lors de l'assemblée générale du 31 mai 2012 le changement de nom (de SPECIF à

1. Professeure des universités, université de Toulon.

DOI:10.48556/SIF.1024.20.31

SIF), une modification de son statut en société savante et a donné un avis favorable à la demande de labellisation « ARUP ».

Plusieurs étapes et plusieurs années ont été nécessaires pour constituer un dossier, enfin complet, à présenter au conseil d'État. La bande dessinée qui suit illustre ce long parcours du combattant, combattant parfois joyeux, parfois usé, parfois enfin dans une attente qui n'en finit pas. Cette reconnaissance d'utilité publique, tant attendue, a été accordée en 2018 !

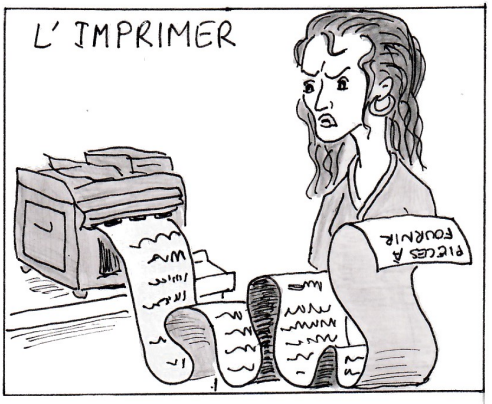
En 2020, la crise sanitaire a montré que les statuts de la SIF devaient encore être retouchés pour se conformer aux usages fréquents de visioconférence et permettre le vote à distance. Un nouveau dossier permettra en 2021 de procéder à ces modifications en conformité avec la loi et d'obtenir une nouvelle fois un avis favorable du conseil d'État !

Je remercie chaleureusement Frédéric Havet pour avoir si bien retranscrit cette odyssée par son magnifique coup de crayon !

« *Le chemin est long du projet à la chose* » disait Molière (Tartuffe)


ÉLISABETH, CE SERAIT BIEN QUE LA SIF SOIT RECONNUE D'UTILITÉ PUBLIQUE. TU VEUX BIEN TE CHARGER DU DOSSIER ?
C'EST COMME SI C'ÉTAIT FAIT!
SIF

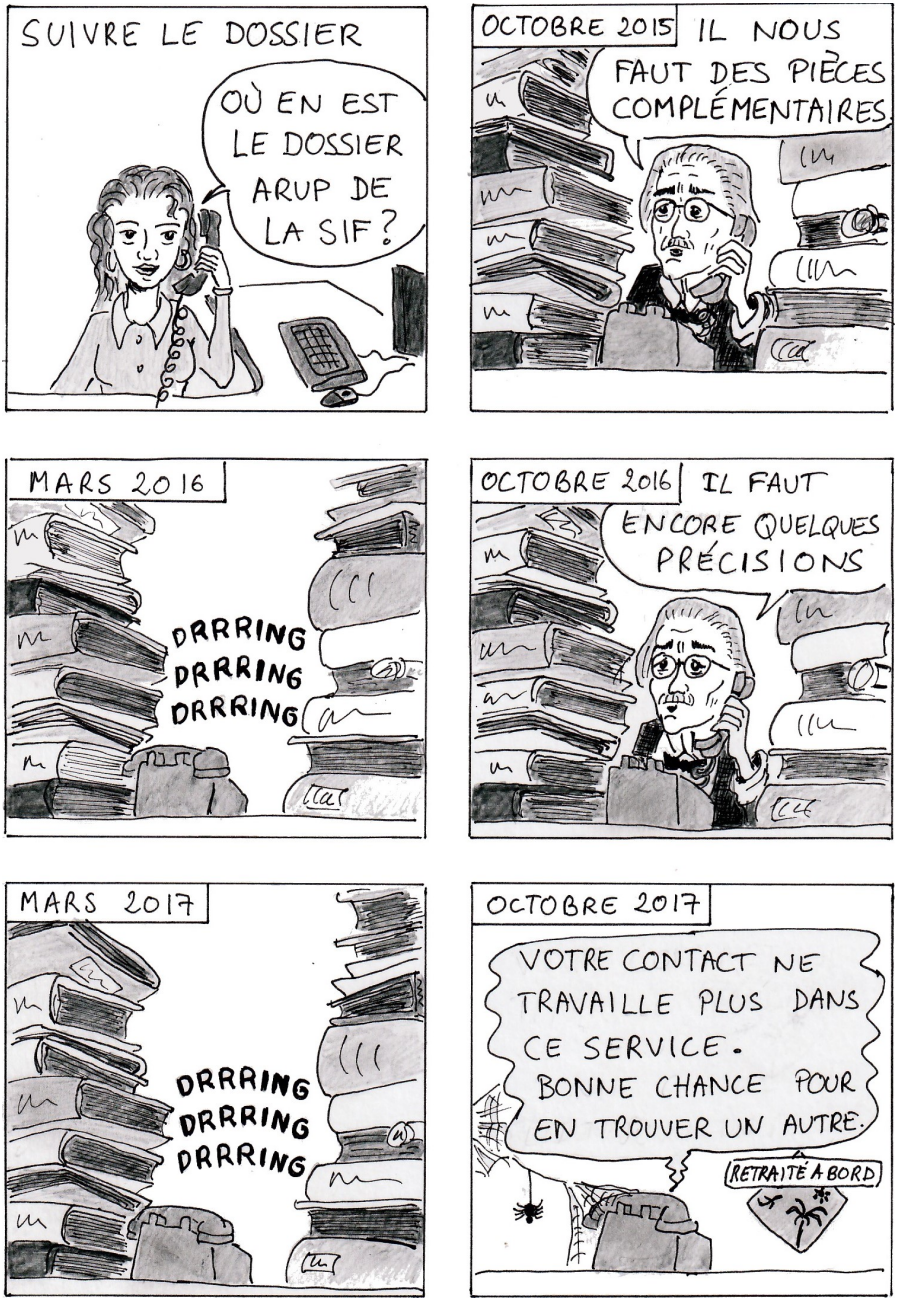

TÉLÉCHARGER LE DOSSIER

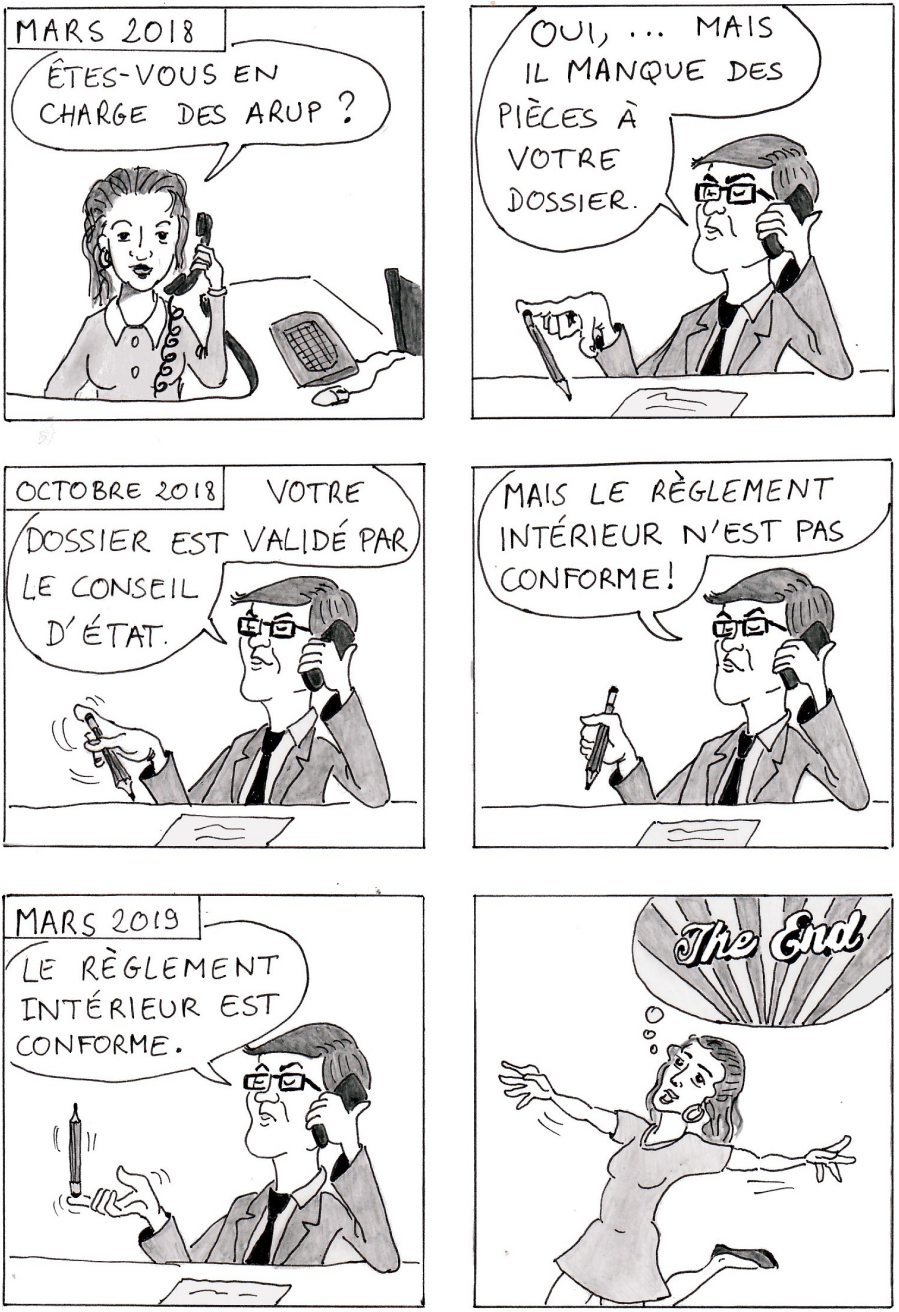

L'IMPRIMER
PIÈCES À FOURNIR


RÉCUPÉRER LES ARCHIVES
SPECIF


S'Y PLONGER


ADAPTER LES STATUTS
SIF
SIF
INTERROGER LES ANCIENS PRÉSIDENTS
STATUTS SIF
ARUP-SIF
REFAIRE LES COMPTES
RÉCUPÉRER LES LISTES D'ADHÉRENTS


SUIVRE LE DOSSIER
OÙ EN EST LE DOSSIER ARUP DE LA SIF?
OCTOBRE 2015
IL NOUS FAUT DES PIÈCES COMPLÉMENTAIRES
MARS 2016
DRRRING DRRRING DRRRING
OCTOBRE 2016
IL FAUT ENCORE QUELQUES PRÉCISIONS
MARS 2017
DRRRING DRRRING DRRRING
OCTOBRE 2017
VOTRE CONTACT NE TRAVAILLE PLUS DANS CE SERVICE. BONNE CHANCE POUR EN TROUVER UN AUTRE.
RETRAITÉ À BORD


MARS 2018
ÊTES-VOUS EN CHARGE DES ARUP ?
OUI, ... MAIS IL MANQUE DES PIÈCES À VOTRE DOSSIER.
OCTOBRE 2018
VOTRE DOSSIER EST VALIDÉ PAR LE CONSEIL D'ÉTAT.
MAIS LE RÈGLEMENT INTÉRIEUR N'EST PAS CONFORME!
MARS 2019
LE RÈGLEMENT INTÉRIEUR EST CONFORME.
The End


MAI 2020
PEUT-ON FAIRE LES A.G. ET LES C.A. EN VISIO ?
TOUT À FAIT MAIS ALORS...
...IL FAUT REFAIRE LE DOSSIER ARUP
TOUT REFAIRE?
OUI POUR ÊTRE CONFORME.
TOUT OU JUSTE LES STATUTS ?
LES STATUTS, LE RÈGLEMENT ET LES AUTRES DOCUMENTS
OCTOBRE 2020
DÉPÔT
RÈGLEMENT INTÉRIEUR
STATUS
DOCUMENTS
MARS 2021
SIF : ARUP
APPROUVÉ
THE END (?)
F.H.

Table ronde sur la recherche en informatique

Serge Abiteboul [1]

À l'occasion du dernier congrès de la Société informatique de France, on m'a demandé d'animer une table ronde dédiée aux principales avancées scientifiques en informatique au cours des dix dernières années, ainsi qu'aux défis des dix prochaines années. Le domaine progresse très vite et il est important d'interroger cette progression. J'ai réuni pour parler du sujet une véritable *dream team*. Je leur ai laissé une liberté complète. Dans un premier temps, ils regardaient le passé, et dans un second, le futur. Précision importante, je ne leur ai pas demandé de réaliser un podium. Il ne s'agit pas de la cérémonie des César, ou de quelque classement que ce soit. Ce ne sont que cinq points de vue, mais ils sont intéressants parce qu'ils viennent de cinq *leaders* de leurs domaines.

Les dix dernières années

Véronique Cortier, directrice de recherche CNRS, Loria

J'ai noté deux changements importants dans le domaine de la sécurité informatique, mon domaine actuel de recherche. Le premier est technique. Il y a 20 ans naissaient les premiers outils basés sur les méthodes formelles (logique du premier ordre, réécriture, etc.) pour analyser les protocoles cryptographiques, qui régissent les échanges sécurisés. Nous étions fiers de les appliquer sur de petits exemples-jouets, sortis de leur contexte. Ces outils ont acquis une grande maturité ces dix

1. Directeur de recherche Inria, Paris.

DOI:10.48556/SIF.1024.20.39 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 39–44

dernières années et sont désormais utilisés lors de la conception des nouvelles générations de protocoles comme TLS ou Signal. Un deuxième point est plus sociétal. La sécurité informatique peut tout autant protéger les citoyens et les institutions contre des attaques, protéger la vie privée, ou au contraire permettre une certaine forme de surveillance (pour la santé ou la sécurité des personnes par exemple). Le cas très médiatisé des applications de traçage en est un exemple. Définir un compromis acceptable entre surveillance et respect de la vie privée va au-delà du rôle de l'informaticien et implique de nombreuses considérations éthiques.

Serge Fdida, professeur, Sorbonne Université

L'histoire des technologies télécom est instructive à bien des égards, et questionne le bon niveau d'abstraction des objets à manipuler dans un réseau. En 1909, avec l'apparition de la téléphonie, le niveau d'abstraction est celui du circuit sur lequel vont s'écouler nos communications téléphoniques. En 1970, nous observons l'émergence de l'Internet suivant une architecture en rupture qui s'appuie sur le fameux paquet IP. Puis, on réalise qu'un certain nombre d'objets partagent des éléments communs et nous parlerons de flots, un ensemble de paquets IP liés par des adresses identiques, par exemple un flot TCP. Mais nous devons encore nous rapprocher de l'application qui fait sens en termes de service offert au client, nous parlerons alors de contenu, ce qui provoque le déploiement accéléré des CDN (réseaux de distribution de contenu). Enfin, pour généraliser les objets numériques manipulés, le concept de programme représente le meilleur candidat permettant de traiter une majorité des fonctions réseaux. Nous sommes donc passés en quelques années d'une architecture distribuée fondée sur internet, mais reposant sur des équipements propriétaires vers une architecture plus centralisée, reposant sur le *cloud*, et dont les fonctions pourront être déployées de façon continue et intégrées dans l'architecture grâce à la flexibilité du logiciel, en s'appuyant sur du matériel de commodité.

Nicholas Ayache, directeur de recherche Inria, Sophia Antipolis

En imagerie médicale et médecine numérique, on a observé ces 10 dernières années une percée spectaculaire des algorithmes d'apprentissage profond, notamment des réseaux de neurones à convolution, des réseaux génératifs antagonistes (GAN), des auto-encodeurs variationnels, etc. Ces algorithmes ont permis des avancées majeures en reconstruction d'images (réduction du bruit, meilleure résolution, accélération des calculs, etc.) et en analyse d'images que ce soit pour la détection, localisation et caractérisation d'anomalies (cancers, etc.) ou pour le recalage (temporel, entre sujets) et la fusion d'images (entre modalités) ainsi qu'en synthèse d'images (prédire une modalité à partir d'une autre). Enfin, ces algorithmes permettent une intégration efficace d'informations très hétérogènes sur le patient en combinant par exemple des images anatomiques et fonctionnelles avec des informations cliniques, biologiques, génétiques, voire comportementales et environnementales (exposome). Beaucoup de

ces avancées se traduisent progressivement par de nouveaux logiciels exploitables en clinique après leur validation par les organismes compétents.

Sonia Ben Mokhtar, directrice de recherche CNRS, LIRIS

Durant les dix dernières années, l'ubiquité de l'informatique est devenue une réalité. Nous sommes en effet submergés de dispositifs informatiques communicants (téléphones, tablettes, montres connectées et de plus en plus d'ordinateurs sont embarqués dans nos objets du quotidien). Ceci est dû à des avancées matérielles majeures : développement de processeurs de plus en plus petits et puissants et de moins en moins chers, connectés à travers des réseaux de plus en plus rapides. En plus de permettre d'aller vers de l'extrêmement petit (informatique embarquée, Internet des objets), ces évolutions matérielles ont permis de faire émerger des infrastructures de stockage et de calcul de plus en plus grandes (centres de calculs, informatique dans les nuages) menant ainsi à des évolutions logicielles majeures (logiciels qui s'appuient sur ces infrastructures tels que les réseaux sociaux, les services géo-localisés, les systèmes de recommandation). Ces logiciels sont également devenus ubiquitaires et s'appuient massivement sur la collecte et l'analyse de données personnelles.

Claire Mathieu, directrice de recherche CNRS, IRIF

Dans le domaine de la conception d'algorithmes polynomiaux qui résolvent des problèmes d'optimisation avec un facteur d'approximation, le problème du voyageur de commerce est un problème phare. La meilleure heuristique connue auparavant, celle de Christofides, avait un facteur d'approximation de 1.5. L'an dernier a été publié la première légère amélioration depuis 50 ans, dans un article de Nathan Klein, Anna Karlin et Shayan Oveis Gharan. Au lieu de commencer par construire un arbre couvrant minimum, ils construisent un arbre aléatoire tiré d'une certaine distribution définie à partir de la relaxation linéaire de Held-Karp pour le voyageur de commerce. Cela permet d'éviter les situations de pire cas pour l'heuristique. Ce résultat montre



une sophistication croissante du domaine qui lui permet, en combinant probabilités, programmation mathématique et outils combinatoires, de progresser enfin dans la compréhension de problèmes qui forment le socle de notre domaine.

Les dix prochaines années

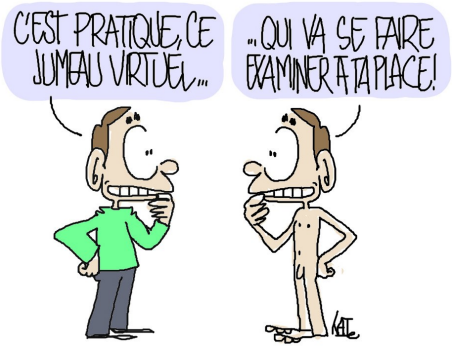
Véronique Cortier

Je pense qu'un enjeu non scientifique nous attend pour les dix prochaines années, à savoir le système de sélection des publications. Le nombre de soumissions explose et avec lui, le nombre de *reviews*, de discussions, la taille des comités de programme. Les conférences jouent un rôle important, à savoir présélectionner les articles à lire dans chaque domaine. Elles sont également un outil central d'évaluation, du recrutement du jeune chercheur à sa promotion tout au long de sa carrière. Cet aspect est plus pervers puisqu'il pousse à une course à la publication et à présenter ses résultats sous le meilleur angle possible. Ces dernières années ont déjà vu des changements importants avec le passage de plus en plus fréquent à l'anonymisation des soumissions (*blind* ou *double blind*), la multiplication des cycles de soumissions (possibilité de soumettre trois fois dans l'année pour une même conférence). La crise Covid ainsi que, de façon plus profonde, la crise écologique obligent à revoir le mode de fonctionnement des conférences : est-il soutenable de se rendre à cinq ou six conférences chaque année, dans le monde entier ? Les 10 prochaines années nous amèneront certainement à revoir notre façon de publier.

Serge Fdida

L'évolution du logiciel est transformante, tant sur la méthodologie, que les compétences, la conception ou les opérations d'une infrastructure numérique qui intègre désormais des objets variés du *cloud* à l'*Internet of Things* (IoT). Le réseau devient une plate-forme virtualisée programmable qui exploite des concepts et des outils utilisés dans ce cadre. L'impact est énorme dans la mesure où l'objectif est de déconstruire (on parle de désagrégation) un routeur, un réseau d'accès radio (RAN) ou l'ensemble d'un réseau pour ensuite le reconstruire en utilisant si possible des briques logicielles open-source déployées à façon. Ce principe s'inspire de celui des PCs s'appuyant sur une base matérielle (*chipset* et *hardware*) et un OS sur lequel s'exécutent des programmes indépendants, ou encore ceux de l'Appstore/Android Market tournant sur un OS ad hoc et un environnement matériel. L'impact industriel qui s'ensuit sera certainement important. Cette évolution illustre l'impact croissant du logiciel dans la conception des systèmes en raison de leur évolution rapide et de leur flexibilité. Par ailleurs, l'ouverture de ces systèmes devient un terreau pour l'innovation. Cette transformation illustrée ici dans le monde des télécoms est observable dans de nombreux autres domaines.

Nicholas Ayache



Les 10 prochaines années devraient voir le développement généralisé de la médecine numérique sur plusieurs fronts. D'abord, pour soulager les médecins spécialistes, dans le cadre des dépistages à grande échelle, une première lecture automatique des images et données du patient pour réduire la masse des dossiers à analyser plus finement ; par ailleurs, des outils d'aide au diagnostic exploitant des appareils portés par les patients eux-mêmes ou opérés par des non-spécialistes (techniciens, généralistes) pour guider l'orientation vers des spécialistes. Ensuite, le développement généralisé du jumeau numérique du patient pour guider le diagnostic, le pronostic et la thérapie, notamment grâce à des outils de simulation numérique. Ceci nécessitera de compléter les modèles statistiques d'apprentissage profond (entraînés sur des bases de données mondiales et multicentriques, centralisées ou distribuées) par des modèles biophysiques de l'anatomie et de la physiologie du patient pour mieux interpréter, prédire et comprendre les raisons d'un dysfonctionnement et les effets attendus d'une thérapie.

Sonia Ben Mokhtar

L'extrême centralisation des services en ligne d'aujourd'hui ainsi que les nombreuses fuites de données qu'ils subissent de manière régulière forcent le questionnement sur la préservation de la vie privée de leurs usagers. Aussi bien le législateur (commission européenne avec des textes de loi tels que la RGPD) que le monde technologique (chercheurs et entreprises) travaillent sur cette problématique. Une piste prometteuse pour les années à venir consiste en la décentralisation (d'une partie) du Web. Plusieurs initiatives ont débuté depuis l'apparition des cryptomonnaies et plus généralement de la Blockchain (par exemple la finance décentralisée), mais d'autres travaux qui ne seront pas forcément liés à la Blockchain vont probablement voir le jour. Je pense en particulier à des travaux autour de la décentralisation des algorithmes d'apprentissage qui sont au cœur de nombreux services en ligne. Décentraliser les algorithmes d'apprentissage soulève néanmoins de nombreux défis et nécessitera certainement la collaboration entre chercheurs de différentes communautés (apprentissage, optimisation, réseaux, algorithmique et systèmes distribués, sécurité et vie privée).

Claire Mathieu

De plus en plus, les rouages de la société tournent grâce au numérique et aux algorithmes. Parcoursup a généralisé APB à presque toutes les formations de l'enseignement supérieur en France. Waze, Google maps et les applications de calcul de plus court chemin, avec évolution dynamique au fil des embouteillages et des accidents, ont révolutionné la façon dont les usagers appréhendent les itinéraires. L'algorithmique peut-elle aider au découpage de circonscriptions en France ? À une rationalisation des micro-règles du calcul de l'impôt sur le revenu ? Comment incorporer les aspects algorithmiques dans les discussions entre les divers partenaires du débat public ? Ces questions guideront une partie de la recherche en algorithmique pour des problèmes « sociétaux ».

Discours de présentation du blog Binaire

Serge Abiteboul [1]



Le blog Binaire est un blog invité du journal Le Monde depuis 2014. Il est intéressant de raconter comment cela a démarré, parce qu'en fait, ça a démarré avec la SIF. À l'époque, je venais d'être nommé président du conseil scientifique de la SIF, et un copain, David Larousserie, journaliste au Monde, est venu me voir en me disant : « *On ne parle pas assez de la science informatique au Monde : est-ce que tu n'as pas envie de faire un blog invité sur ce thème* ? » J'ai eu envie de répondre : « *Non, merci, j'ai déjà trop de choses à faire, je ne peux pas* ». Mais, je me suis dit : « *Moi non, mais peut-être que le conseil scientifique de la SIF, oui ?* ». En fait, cela ne s'est pas fait avec le conseil scientifique de la SIF, mais avec un groupe d'amis. Vous en connaissez sûrement certains, comme Thierry Viéville, Marie-Agnès Enard, Pierre Paradinas, Charlotte Truchet, qui était là hier, et d'autres. Je vais rater des gens si j'essaie de lister tout le monde, donc je vais m'arrêter au noyau initial. Donc, une bande de copains décide de réaliser ce blog ; on assume : on n'a pas d'organisation, on est assez bordéliques, mais on fonctionne.

Nos rapports avec Le Monde étaient très bons avec les journalistes, moins avec l'administration, parce qu'on a décidé de manière autoritaire que le journal serait en

1. Directeur de recherche Inria et membre de l'Académie des sciences.

DOI:10.48556/SIF.1024.20.45 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 45–50

licence Creative Commons [2]. Tous les articles peuvent donc être reproduits par qui veut à condition de nous citer. Le Monde est gentiment venu me dire : « *On vous donne 50 % des royalties sur la publicité, signez, en bas de cette page* ! » Mais, dans le contrat, il était marqué que j'abandonnais la propriété des articles au Monde, alors j'ai dit : « *Je ne peux pas vous abandonner quelque chose que je ne possède pas, c'est en Creative Commons* ». Ils m'ont dit : « *C'est quoi ça ? On va voir avec le service juridique* ». Au bout de quatre ans, ils sont revenus me voir en me disant : « *Mais comment ça se fait que vous n'ayez toujours pas signé le contrat ?* ». J'ai proposé de signer le contrat sauf quelques articles que je ne pouvais pas signer. Ils m'ont dit : « *On va voir avec les services juridiques* ». On en est là. Le bon côté, c'est qu'on est toujours en Creative Commons, le côté négatif, c'est qu'on n'a jamais touché un centime.

Notre sujet, c'est l'informatique. Je vais vous la lire notre mission : « *L'informatique participe au changement profond du monde dans lequel nous vivons, mais qu'est-ce que l'informatique, quels sont ses progrès, ses dangers, ses questionnements, ses impacts, ses enjeux, ses métiers et son enseignement* ? » C'est à cela qu'on répond. On a créé des liens. Avec)I(nterstices [3] (on est des fans d'Interstices), de temps en temps, on partage des articles avec eux. Avec 1024 évidemment, et avec la médiation dans la fondation Blaise Pascal... Il existe tout un écosystème sur la médiation en France dont on tient à être partie avec notre particularité qui est d'être un blog grand public. Notre but dans la vie, c'est que les lecteurs du Monde nous lisent régulièrement, tous les lecteurs du Monde.

Vous allez me demander : « *Combien de lecteurs du Monde vous lisent* ? » Pas tous. Ça se compte en quelques milliers pour l'article de base. Pour un article plus flashy comme « Les algorithmes et la sexualité », vous passez à plusieurs dizaines de milliers. C'est un très bel article de Christos Papadimitriou. On est aussi repris ; le fait d'être Creative Commons, ça a un avantage, c'est que les gens peuvent vous reprendre. Il y a des journaux avec qui on a des affinités, comme The Conversation [4], que vous connaissez, qui nous reprennent régulièrement, et puis il y en a d'autres qui, des fois, nous reprennent sans nous le dire. De temps en temps, vous découvrez que vous avez été repris dans un journal *X* ou *Y*. Bon, nous, on aime ça, parce que le but, ce n'est pas notre ego, c'est qu'on parle d'informatique.

Alors, de quoi parle-t-on ? Notre sujet fétiche, ce n'est pas surprenant, c'est la science et les techniques informatiques. Et puis, la médiation, notre but, c'est de faire comprendre ce que c'est qu'un vote électronique, ce que c'est que l'IA, ce que c'est que la 5G... Tous ces trucs-là, les expliquer dans des termes que tout le monde comprend. Régulièrement, Marie-Agnès Enard revient en nous disant : « *Cet article, tu n'y es pas du tout, personne ne va comprendre* ». Elle nous engueule mais c'est

2. https://creativecommons.org/licenses/?lang=fr-FR.
3. https://interstices.info.
4. https://theconversation.com/fr.

comme ça qu'on fonctionne. On a des rubriques culte, comme « *Il était une fois ma thèse* », qui reprend en particulier les prix de thèses de Gilles Kahn où on demande aux étudiants d'écrire ; « les Entretiens de Binaire », dont je vais reparler...

Et je veux finir cette première partie sur Binaire en disant qu'essentiellement, on est une émanation de la SIF donc n'hésitez pas à nous aider. Le conseil scientifique fait beaucoup ; on a des gens qui publient régulièrement dans Binaire : Claire Mathieu, Anne-Marie Kermarrec... Nous sommes aussi des éditeurs, les petites mains. Quand on n'écrit pas, on prend les articles des autres et on les met sur le logiciel du Monde... Un logiciel qui n'est pas terrible. Il y a le travail de réécriture, de discussion avec les auteurs. Si vous avez des idées d'articles, allez-y, écrivez-nous, proposez-nous des sujets, écrivez des articles. Une autre façon de contribuer qu'on aime bien, c'est de dénoncer vos collègues : « *Je partage le bureau avec Machine et elle a fait un truc super, elle aimerait bien en parler mais elle n'ose pas, elle hésite*. » Balancez-la !

On aimerait bien toucher plus de monde... Alors, je vais faire des vœux pieux : on aimerait bien avoir un public plus jeune, on ne sait pas comment faire, donc on n'hésite pas à inviter des étudiants, des lycéens à écrire. C'est rare, on ne sait pas comment les toucher. Si vous avez un moyen de nous faire rajeunir, venez. On aimerait bien aussi toucher plus de profs de collège et de lycée, une manne de personnes au contact avec beaucoup de monde, en contact avec la vraie vie. On est déçus qu'ils n'écrivent pas plus. Donc si vous pouvez nous relayer...

Et puis, il y a des choses qu'on aimerait bien faire depuis longtemps et qu'on fait de façon homéopathique comme des *podcasts*. Si vous avez envie de faire un *podcast* régulier, venez, rejoignez-nous, lâchez-vous. Si vous voulez faire des vidéos, pourquoi pas ?

Ça, c'était le premier sujet, Binaire, j'espère que vous avez tous compris qu'on est extrêmement demandeurs du soutien franc et massif de la communauté SIF pour écrire dans Binaire et, évidemment, pour nous lire, pour nous relayer sur les réseaux sociaux, nous faire connaître. Allez-y, faites ça, on a besoin de votre soutien !

Les entretiens de Binaire...

Les Entretiens de Binaire ont commencé pratiquement au tout début de Binaire. Le conseil scientifique a écrit un texte qui était « L'informatique : la science au cœur du numérique ». On était contents de notre texte, on l'a publié sur Binaire et on aurait pu s'arrêter là. Mais, en relisant le texte, on s'est aperçus que ce n'était pas que cela, l'informatique. C'était super bien, mais c'était l'informatique décrite de l'intérieur, comme les informaticiens la voient. C'était très large, très général, avec des frontières qui englobaient la robotique, le traitement du signal... On dirait aujourd'hui : la « science du numérique », pour faire moderne, même si je préfère le mot informatique. Et puis on s'est dit : « L'informatique, c'est aussi ses liens avec les autres

disciplines », ce dont a parlé Antoine Petit. On s'est demandé comment raconter ce lien avec les autres disciplines. Et on a lancé les entretiens autour de l'informatique. Cela consiste à s'entretenir avec des gens qui travaillent avec l'informatique mais qui ne sont pas forcément des informaticiens : une chimiste, un biologiste, un physicien, une philosophe, des gens qui connaissent bien l'informatique mais qui ne sont pas informaticiens. On les interviewe et on leur demande de raconter les liens entre ce qu'ils font et l'informatique. Donc c'est une description un peu impressionniste de l'informatique, au lieu d'être de l'intérieur, on est un peu loin, on fait ça par petites touches, et ces touches, il y en a une quarantaine maintenant, depuis 2014. C'est des touches qui, à mon avis, décrivent l'informatique autrement mais de façon au moins aussi intéressante que le texte qu'on avait écrit entre informaticiens. Par exemple, il y a un entretien avec François Houllier, qui était le président de l'INRA, qui nous raconte ce que c'est que l'informatique en agronomie, et moi, je trouve ça passionnant. Voilà, les Entretiens de Binaire.

Et ensuite...

Que s'est-il passé ensuite ? Un jour, Hervé Le Crosnier qui a été professeur d'informatique dans une vie antérieure, est venu dîner à la maison ; on papotait, on avait peut-être picolé un peu trop, et il m'a dit : « *Dans ma maison d'édition, C&F Éditions*[5]*, je te propose de publier des Entretiens de Binaire* ». J'ai répondu : « *Super idée, ça me fait plaisir parce que je suis fana de ces entretiens, ça va leur donner une deuxième vie, ça va les faire lire par un autre public* ». Mais je me suis aussi dit : « *Une fois qu'il sera rentré chez lui, il aura oublié* ».

C'est vrai qu'il ne s'est rien passé pendant un certain temps. Et puis, je vais faire un exposé à Caen, et je tombe sur une maître de conférences, Elsa Jaubert, qui me dit : « Hervé m'a demandé de regarder les entretiens pour en publier dans un bouquin ». Et puis, il se passe encore du temps et la COVID, et Hervé me présente Coralie Mondissa qui était en train d'éditer le bouquin. C'était devenu une réalité. Donc, la bonne nouvelle, la voici : un certain nombre d'entretiens autour de l'informatique, une quinzaine, vont être publiés par C&F Éditions, une maison d'édition citoyenne.

Je vais quand même expliquer ce que ça veut dire « une maison d'édition citoyenne ». D'abord, en voyant la liste de gens qu'ils publient, vous allez peut-être comprendre un petit peu : ils publient des gens comme Stéphane Bortzmeyer, Danah Boyd, Tristan Nitot, ou Fred Turner. Bon, ils ne publient pas Mark Zuckerberg. Leur idée, c'est de parler de culture numérique, et pas pour dire : « C'est génial, la culture numérique », mais pour en parler d'un point de vue critique, et militant. Ils ont une certaine façon de prendre des valeurs et de les défendre. Nous, ça nous va bien parce que les valeurs qu'ils défendent, ont un *overlap* énorme avec les valeurs que défend Binaire. Pour ceux qui lisent Binaire, vous avez dû comprendre que la

5. https://cfeditions.com.

science, pour nous, c'est une valeur qu'il faut défendre, la culture scientifique, c'est une valeur qu'il faut défendre, qui est un peu, de temps en temps, mise sous le tapis, et c'est quelque chose, avec la médiation, qu'il fait vraiment pousser, mais il y a d'autres valeurs, comme la parité : on a énormément d'articles là-dessus, les liens entre démocratie et numérique, etc.

L'informatique doit servir à améliorer le monde dans lequel on vit, et ça, c'est une valeur que Binaire partage avec C&F Éditions, et c'est pour ça qu'on est très contents d'avoir ce bouquin qui paraît chez eux. Mais, la vraie bonne nouvelle, c'est que la SIF, dans sa grande générosité, offre un livre à chacune des personnes dans la salle. Merci la SIF. Est-ce que vous avez des questions ?

> Vous avez le droit de poser une question sur Binaire à la seule condition que vous vous engagiez à rédiger un article ou à dénoncer quelqu'un qui ferait un article. C'est comme cela que ça marche dans Binaire ! On est les rois de la délation.



Participant au congrès : « *Juste une question : Binaire, ça existe, si j'ai bien compris, depuis huit ans, et j'aimerais savoir s'il y a un, deux ou trois articles dont tu te souviens, qui ont été mémorables, qui t'ont marqué ?* »

Serge Abiteboul : C'est une question très difficile. J'aurais tendance à parler des rubriques, parce que ce sont des lieux ou des rencontres régulières. On est très fiers de la rubrique sur les thèses. À titre personnel, ce n'est pas pour vendre ma sauce, j'ai pris un plaisir considérable à faire les Entretiens de Binaire. Il y a d'autres rubriques,

comme Petit Binaire, qui explique aux enfants ce que sont certaines notions d'informatique, donc il y a un grand nombre d'articles que j'adore dans les articles de Binaire. Mais revenons sur les entretiens... Ces entretiens, j'ai quand même baigné dedans depuis plusieurs années, et je les ai vécus avec beaucoup de plaisir, mais le regard de C&F Éditions dessus était tout à fait autre. Moi, ce que je voyais là-dedans, c'était comment expliquer ce que c'est que l'informatique avec ce côté impressionniste. La façon dont Hervé l'a lu est complètement différente : « Ces gens-là ont une passion et ils partagent leur passion de l'informatique ». Cela m'a fait porter un autre regard sur ces entretiens. C'est vrai que c'est ce qui transparait, c'est la passion. Par exemple, Arshia Cont parle de musique, mais aussi d'informatique qu'il découvre et qui le passionne. En fait, ce qui est vraiment très visible dans ces articles, c'est la passion que ces gens ont pour l'informatique et pour leur discipline, que ce soit la chimie, la biologie ou autre. Les scientifiques sont des gens passionnés, c'est ce qu'on voit bien, ils ont la passion de la science. Donc finalement ce qui est mémorable, ce qui m'a marqué, c'est leur passion.

Voilà, c'est ce que je voulais raconter, je vous remercie de votre attention.

Mille vingt-quatre

Denis Pallez et Éric Sopena [1]

Lors du CA du 21 mars 2012 de l'association Specif, alors que le numéro d'avril du bulletin était en passe d'être bouclé, une commission de réflexion composée de Christophe Fiorio, François Jacquenet et Éric Sopena fut constituée pour étudier l'évolution de la transformation du bulletin des personnels en un bulletin d'une société savante. Alors que la Société informatique de France se crée en mai 2012 présidée par Colin de la Higuera, ce dernier propose à Éric Sopena, avec une conviction jusque là inégalée, de prendre les rênes du futur bulletin.

Après discussion en CA de la SIF, il fût décidé de construire un comité éditorial via un appel à volontaires au sein de la communauté pour s'impliquer dans ce nouveau projet. Les premières discussions sur le bulletin de la SIF datent de février 2013 et l'une des premières préoccupations a été de lui trouver un nom. Les idées, plus ou moins sérieuses, ont fusé : Bulletin de la SIF, Bulletin de la Société informatique de France, Les échos de l'informatique, Avenir numérique, Le journal de l'informatique, L'informatique – du primaire au supérieur, Le courrier informatique, NRI (Nouvelles, recherches et informatique), SIF MAG, SIF NEWS, @BuSIF, PENSIF (Publications et nouvelles de la SIF), INCISIF (Informations et citations de la SIF), loGIN SIF, loGIN (Gazette des informaticiens), RES-SIF (Revue scientifique de la SIF), JOUIS-SIF (Journal d'information scientifique de la SIF), EXPLO-SIF (Explorations de la science informatique), 1024, 4096, COMPUTIUS, SIFonaptera, La puce. C'est ainsi que le 2 mai 2013, *1024* l'emporte largement devant SIF NEWS, ex æquo avec le bulletin de la SIF.

1. Enseignants-chercheurs et rédacteurs en chef du bulletin 1024.

DOI:10.48556/SIF.1024.20.51 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 51–53

Nos amis de la Société mathématique de France (SMF) nous ont gracieusement fourni le style LaTeX qu'ils utilisaient pour la Gazette de la SMF [2] afin que nous puissions l'adapter à nos besoins. C'est Thierry Murgue, stéphanois, qui a réalisé cette adaptation. La charte graphique du bulletin, quant à elle, fut conçue par Lolligraph.

Pendant ce temps, Éric Sopena se chargea de référencer le bulletin en demandant l'attribution d'un *International Standard Serial Number* (ISSN), ce qui oblige un envoi de chaque numéro à la Bibliothèque nationale de France. La structure étant définie, il ne manquait plus qu'à la faire vivre, et c'est en faisant appel aux experts du domaine et à des médias partenaires comme)i(nterstices ou TSI que le premier numéro vit le jour en septembre 2013.



COMITÉ DE RÉDACTION

Sylvie Alayrangues
Université de Poitiers
Vincent Autefage
Université Bordeaux 1
Olivier Baudon
Université Bordeaux 1
Colin de la Higuera
Université de Nantes
Jean-Paul Delahaye
Université Lille 1
Olivia Dumas
Académie de Caen
Christine Froidevaux
Université Paris-Sud
Thierry Garcia
Université de Versailles Saint-Quentin-en-Yvelines
Marie-Claude Gaudel
Université Paris-Sud
Claudia Marinica
Université de Cergy-Pontoise
Philippe Marquet
Université Lille 1
Denis Pallez
Université Nice Sophia Antipolis
Vincent Ribaud
Université de Bretagne Occidentale
Éric Sopena
Université Bordeaux 1, rédacteur en chef

Contact : 1024@societe-informatique-de-france.fr

Cette œuvre est mise à disposition sous licence Attribution - Pas de Modification 3.0 France. Pour voir une copie de cette licence, visitez
http://creativecommons.org/licenses/by-nd/3.0/fr/
ou écrivez à Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Société informatique de France
Institut Henri Poincaré, 11 rue Pierre et Marie Curie, 75231 Paris Cedex 05
Directeur de la publication : Colin de la Higuera — ISSN : en cours d'attribution.
Couverture : d'après une maquette réalisée par Lollygraph.com.

Douze numéros plus tard, Éric Sopena quitte le CA de la SIF et confie les rênes du bulletin à Denis Pallez en janvier 2019 pour aller vers d'autres horizons.

À ce jour et après neuf ans d'existence, 1024, c'est 22 numéros dont deux hors-séries, 408 articles écrits répartis en 30 rubriques, imprimés sur 3 190 pages et distribué en 26 650 exemplaires dans environ 100 unités de recherche ou d'enseignement et 500 adhérents de la SIF.

2. https://smf.emath.fr/les-publications/la-gazette.

Comme le disent si bien nos amis du blog Binaire [3], notre bulletin c'est 1024 raisons de comprendre l'informatique à travers des textes qui s'adressent à toutes celles et ceux pour qui l'informatique est un métier, une passion ou qui souhaitent se tenir au courant des dernières innovations du domaine. Il est également ouvert à toutes celles et ceux qui veulent bien prendre le temps de nous communiquer leur savoir, leur expérience ou leurs réactions du moment ! Et contrairement aux mastodontes de la publication scientifique, le comité de rédaction met en page, relis, corrige orthographiquement et grammaticalement et pousse le vice jusqu'à l'illustration de manière décalée. Alors qu'attendez-vous pour nous soumettre votre expérience ou faire partager votre thématique de recherche ?

Il est plutôt aisé de faire un bilan des 10 dernières années mais, au contraire, beaucoup plus délicat de se projeter sur les 10 prochaines. Récemment, vous avez pu constater une amélioration du référencement avec la création d'un numéro d'identification (DOI [4]). Disposer d'un tel numéro pour chacun des articles de l'ensemble des numéros de 1024 depuis sa création devrait nous faciliter un référencement aux journaux en accès libre (DOAJ [5] ou épi-revues [6] par exemple). Alors, selon vous, que devrions-nous souhaiter de plus ou de mieux au bulletin ? Nous vous invitons à vous exprimer via la liste de discussion de la SIF [7].

3. https://www.lemonde.fr/blog/binaire/2021/04/27/1024-raisons-de-comprendre-linformatique/

4. https://www.doi.org.

5. https://doaj.org.

6. https://www.episciences.org.

7. forum@societe-informatique-de-france.fr.

Prolog, 50 ans d'histoire de l'intelligence artificielle

Laurent Cervoni et Julien Brasseur [1].

À l'occasion des 50 ans de la naissance du langage Prolog, nous proposons, dans cet article, d'évoquer certains des aspects qui, selon nous, en font un langage atypique, et nous discuterons de son apport dans les problèmes d'intelligence artificielle. Pour ce faire, quelques exemples concrets seront proposés et discutés.

Introduction : une brève histoire de Prolog

Le langage Prolog a été présenté pour la première fois à Marseille en 1972 par Alain Colmerauer et son équipe [3]. Prolog, contraction de programmation en logique, est un langage permettant de modéliser le raisonnement en s'appuyant sur la logique d'ordre 1. Il est l'aboutissement de réflexions commencées quelques années auparavant, notamment sur les systèmes-Q, qui sont des mécanismes de règles de transformations de graphes orientés, utilisés, par exemple, pour l'analyse syntaxique de phrases. On peut d'ailleurs noter qu'un des points de départ de ces travaux de recherche était la traduction (français-anglais) et le traitement du langage (analyse et génération de phrases, sur la base de la définition de leur structure grammaticale).

Ce travail, enrichi par les recherches de [7] de l'université d'Edimbourg (puis par les travaux de David H. D. Warren et bien d'autres), a jeté les bases de la programmation logique telle qu'elle existe encore aujourd'hui. La syntaxe des premiers Prolog a alors évolué vers la syntaxe dite d'Edimbourg puis a été normalisé par l'ISO [4]. Dans les années 1980, plusieurs variantes ont vu le jour et d'autres technologies

1. Directeur et chercheurs du Centre de recherche et innovation Talan

DOI:10.48556/SIF.1024.20.55 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 55–63

connexes ou dérivées (comme la programmation par contraintes) sont alors devenus des outils clés des approches symboliques de l'intelligence artificielle.

D'après l'index Tiobe (cf. table 1), 50 ans après sa création, Prolog reste toujours dans la vingtaine des langages utilisés (il était 3^{e} dans les années 1980) et semble même connaître un sensible regain d'intérêt. Son cinquantenaire pourrait être l'occasion de renforcer cette tendance, l'apport de la programmation logique étant particulièrement intéressant en spécifications de modèles, en raisonnement basé sur des règles et en résolution de problèmes avec contraintes.

La communauté Prolog est toujours active et le langage est proposé dans des versions modernisées comme Scryer Prolog (développé principalement en RUST), incluant CPL(B) et CLP(Z), Problog qui incorpore des éléments probabilistes, ou inspire de nouveaux outils tel Logica [2], conçu par Google.

Langage	Juin 2022	2017	2012	2007	2002	1997	1992	1987
Python	1	5	8	7	12	28	-	-
C	2	2	2	2	2	1	1	1
Java	3	1	1	1	1	13	-	-
C++	4	3	3	3	3	2	2	5
C#	5	4	4	8	18	-	-	-
Visual Basic	6	15	-	-	-	-	-	-
JavaScript	7	7	10	9	9	20	-	-
SQL	8	10	-	-	-	-	-	-
Assembleur	9	-	-	-	7	-	-	-
Swift	10	8	6	5	6	-	-	-
Prolog	**20**	**33**	**37**	**27**	**16**	**19**	**15**	**3**
Lisp	32	31	13	15	13	10	6	2
Pascal	270	105	16	20	99	9	3	6

TABLE 1: Extrait de l'index Tiobe répertoriant les langages de programmation les plus populaires (`https://www.tiobe.com/tiobe-index`, juin 2022).

La galaxie des outils et langages autour de la programmation logique (cf.figure 1) s'est enrichie au fil des années, Alain Colmerauer ayant lui-même contribué à la programmation par contraintes dans Prolog IV [2].

2. `https://logica.dev`.

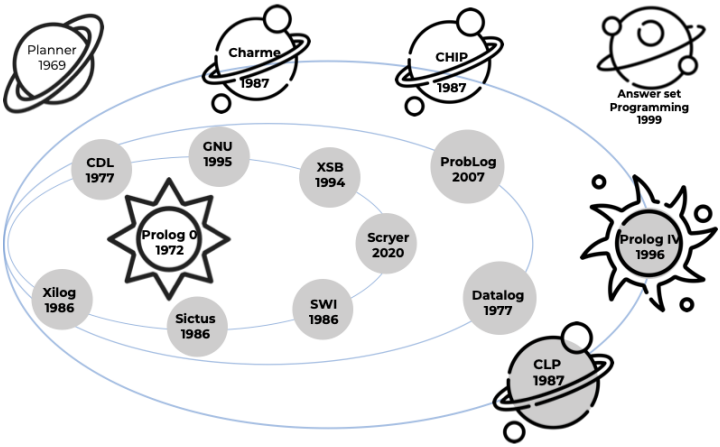


FIGURE 1. Représentation partielle et subjective d'une histoire de la galaxie « programmation logique ».

Le fonctionnement de Prolog et quelques exemples

Dans les langages de programmation « traditionnels » (impératifs), l'investissement pour représenter le raisonnement ou l'expression « humaine » est important puisqu'il faut pratiquement résoudre le problème en le programmant. En Prolog, on reste suffisamment proche de la logique mathématique pour que toute une série de problèmes puissent s'exprimer de façon assez proche de leur simple description.

Prolog utilise ainsi la logique du premier ordre pour permettre la description d'un « univers » en :

— établissant des faits sur des objets ;
— définissant des règles ou des relations sur ces objets ;
— interrogeant le système sur ces objets et ces relations.

Prolog peut aussi, dynamiquement, enrichir sa base de faits et de règles en fonction d'informations nouvelles ce qui rend le langage totalement différent de Python ou d'un autre langage plus courant.

En Prolog, les variables se définissent au sens des variables mathématiques que l'on trouve dans les expressions du type « Quel que soit X tel que... ». Elles ne sont pas typées au sens informatique usuel : une variable Prolog pourra s'unifier avec un entier, une chaîne, une liste ou tout autre objet Prolog. Par convention, toute variable Prolog débute par une majuscule. Ayant établi des relations et un univers, il est évidemment légitime de vouloir le questionner.

Considérons, à titre d'exemple, le programme Prolog de la table 2. Si on pose la question en Prolog de savoir si Socrate est humain (« `?- humain(socrate).` »),

Tout humain est mortel (pour tout X humain, X est mortel)	`mortel(X) :- humain(X).`
Socrate est humain	`humain(socrate).`
Mon voisin est humain	`humain(monvoisin).`

TABLE 2. Expression en français (à gauche) et écriture en Prolog (à droite).

alors la réponse sera logiquement « `true` ». De manière similaire, à la question « `?- humain(Qui).` », Prolog, qui est un langage non-déterministe, propose toutes les réponses valides à une question donnée et retournera donc « `Qui = socrate`, `Qui = monvoisin` ». Cependant, à la question « `?- humain(laurent).` », Prolog retournera « `false` », ce qui est logique puisque cette information ne figure pas dans son domaine de connaissance. On peut évidemment compléter ce « monde » Prolog en l'enrichissant de nouveaux faits et règles (cf.table 3).

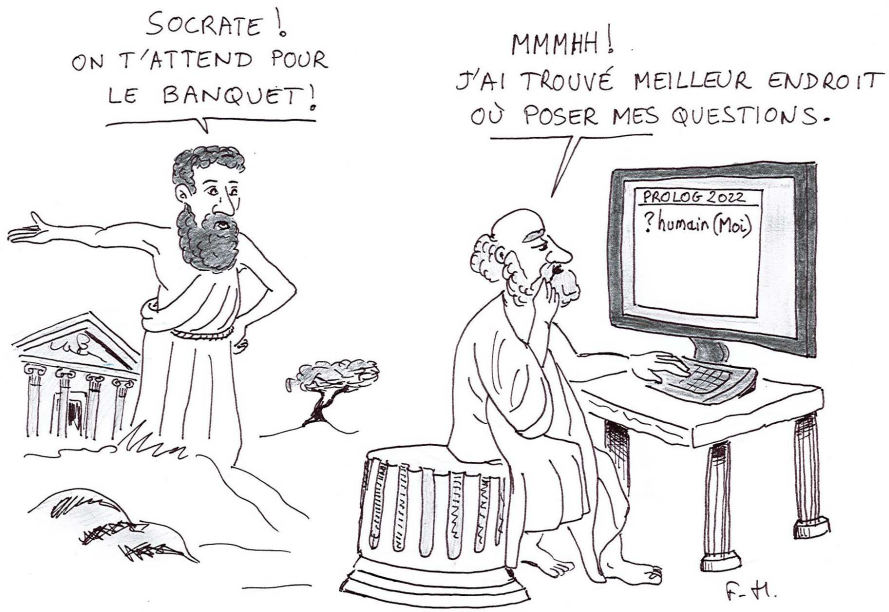
```
mortel(X) :- humain(X).
meurt(X) :- mortel(X), empoisonne(X).
empoisonne(X) :- boit(X,Y), poison(Y).
humain(socrate).
humain(monvoisin).
poison(cigue).
boit(socrate,cigue).
```

TABLE 3. Description plus complexe d'un monde.

Remarquez qu'il est possible de tester les éléments de programme proposés dans cet article dans SWI-Prolog, par exemple dans sa version en ligne [3]. Ces quelques exemples montrent la spécificité de Prolog, qui est donc un langage de programmation atypique. En effet, un programme Prolog est constitué de « paquets » de clauses de Horn. Une clause de Horn est un cas particulier de clause, composée d'une disjonction logique de littéraux, où au maximum un des littéraux est positif, et tous les autres, clos ou non clos (i.e. avec ou sans variable au sens logique), sont négatifs. La tête de la clause est le littéral positif, le corps de la clause contient les autres éléments.

Une clause de Horn « sans corps », c'est-à-dire avec un seul littéral positif est un fait. Un paquet contient les clauses ayant le même foncteur de prédicat (i.e. le même nom de clause). Une clause de Horn peut donc s'écrire sous la forme $(a \wedge b \wedge c \wedge \ldots) \Rightarrow$ Tête. L'exécution d'un programme en Prolog revient à vérifier si un but (ou une requête) est vrai au regard de l'ensemble des clauses contenues dans

3. `https://swish.swi-prolog.org`.



ce programme. Deux mécanismes essentiels sont mis en œuvre lors d'une requête Prolog : l'unification et le principe de résolution.

Unification Le principe de l'unification est le suivant : deux termes t_1 et t_2 sont dits unifiables s'il existe s, une substitution des variables, telle que $s(t_1) = s(t_2)$. Par exemple, si $t_1 = f(X,a)$ et $t_2 = f(b,Y)$, alors t_1 et t_2 sont unifiables pour $X = b$ et $Y = a$. En Prolog, l'algorithme de Herbrand calcule l'unificateur le plus général (*most general unifier* ou *mgu*, i.e. celui qui unifie le moins les variables). Si t_1 et t_2 sont unifiables, l'algorithme calcule le *mgu* des deux termes, sinon il signale un échec : en Prolog standard, il est convenu que t et X sont unifiables si t ne contient pas X.

Résolution Décrivons maintenant l'algorithme de résolution. On appelle « but » une clause de Horn sans littéral positif et « hypothèses » les autres. Pour résoudre un système de clauses de Horn par le principe de résolution, il est nécessaire d'avoir (1) un but (T) et (2) un ensemble d'hypopthèses (C_i).

Supposons que la clause T doive être prouvée à partir des clauses $C_1, \ldots, C_i, \ldots, C_n$, le fonctionnement du processus de résolution peut schématiquement se résumer en quatre étapes :

1. ajouter $\neg T$ à l'ensemble des C_i ;
2. calculer le résolvant de deux clauses ;

3. si le résolvant est vide (antilogie), la preuve est faite ;
4. sinon, l'ajouter à l'ensemble et retourner à l'étape 2.

Il est à noter que la stratégie de Prolog est dite en profondeur d'abord. Elle peut se perdre dans des branches infinies. L'ordre des éléments dans le but et l'ordre des clauses dans les paquets détermine la construction de l'arbre de recherche. L'utilisation de certains prédicats (comme l'opération `cut` qui supprime certaines branches de l'arbre des clauses à parcourir) viennent modifier la stratégie de parcours et réduire la lisibilité d'un programme.

Prolog aujourd'hui — Le cas Recov'Up

Ce panorama sur Prolog ne serait pas complet sans mention de ses applications les plus récentes, dont une liste non exhaustive peut être trouvée dans [6]. Considérons, à ce titre, le cas de l'application Recov'Up, application récente basée sur Prolog.

Recov'up est une application web qui prévient et aide les patients à la rééducation des troubles musculosquelettiques comme l'entorse de cheville, la lombalgie, la cervicalgie, etc. Cette application élabore une liste d'exercices en fonction des symptômes du patient.

La génération des exercices est obtenue par un algorithme d'intelligence artificielle symbolique, développé en Prolog, en collaboration avec les équipes du centre de recherche de Talan et de la startup Skeewai (éditrice de Recov'up). Cette approche symbolique s'appuie principalement sur des règles de rééducation exprimées par les médecins et kinésithérapeutes, spécifiant les exercices pertinents pour le patient en fonction de sa pathologie et son état physique. Le mécanisme mis en œuvre ne nécessite pas un apprentissage sur un important volume de données (ce qui est un avantage dans le secteur de la santé où les règles d'utilisation des données des patients sont strictement encadrées) mais intègre les données spécifiques de chaque patient et son évolution face aux séances qui lui sont proposées.

La génération d'exercices de rééducation s'appuie d'une part, sur un cœur en logique d'ordre 1 « standard » (la description des règles médicales et de kinésithérapie qui permettent de déterminer quels sont les exercices qui correspondent à la situation d'un patient) et, d'autre part, sur la capacité de Prolog à exécuter dynamiquement des prédicats et à enrichir sa base de clauses (via les prédicats d'ordre 2 que sont `findall`, `bagof` et `assert`).

Prolog et l'intelligence artificielle symbolique

L'intelligence artificielle peut être scindée en deux grands courants : l'intelligence artificielle symbolique et l'intelligence artificielle « connexionniste » (ou « numérique »). Cette dernière connaît, depuis 2012 environ, une popularité et une croissance vertigineuse due, entre autres, au développement des technologiques de l'information et à l'acquisition des larges quantités de données qu'elles ont permises.

Toutefois, l'IA numérique peut exiger, pour produire des résultats probants, des quantités de données parfois tellement importantes qu'elles nécessitent, pour être traitées (particulièrement en phase d'apprentissage), une puissance de calcul conséquente, ce qui peut être rédhibitoire, notamment pour les entreprises. En dehors des problèmes pratiques que cela pose, il est à noter que la quantité d'opérations qui doivent être effectuées peut avoir un impact carbone [13] ou une consommation énergétique qu'il est aujourd'hui nécessaire de limiter.

Il existe aussi de nombreux cas pratiques où les données peuvent être indisponibles, impossibles à acquérir ou ne pas pouvoir être conservées suffisamment longtemps (respect du RGPD, volumétrie...). Le cas se pose classiquement lorsque l'on souhaite mettre en place une IA statistique dans un cadre médical, où les données sont sensibles par nature et protégées par un cadre légal strict dès que des données personnelles sont en jeu.

L'IA symbolique et, plus particulièrement Prolog, peut ici être un atout majeur venant compléter la puissance de l'IA numérique. Le cas du projet Recov'Up, cité plus haut, en est un exemple : une pathologie est soignée quotidiennement par les médecins mais il n'existe pas de larges bases de données décrivant l'évolution du patient en fonction des différents soins pratiqués et des profils de patients.

Prolog présente également un intérêt pour les cas où il est important que l'IA soit explicable. Dans certains cas d'applications pratiques, civiles (voitures autonomes, banques, médecine, etc) comme militaires (drones, etc), cela peut devenir critique. En effet, les résultats donnés par Prolog sont, par nature, explicables : « coder » en Prolog revient à décrire un univers, aussi l'explication est-elle *contenue* dans la modélisation Prolog. On peut, en outre, interroger Prolog sur les unifications de faits réalisées : Prolog les parcourt en profondeur et peut expliquer son raisonnement via le système inductif qu'il utilise.

C'est ainsi, d'ailleurs, que Prolog a été utilisé pour concevoir des automates de détections de bugs dans le code de Tensorflow [12] ou encore pour réaliser des démonstrateurs de théorèmes, comme Muscadet [10] ou ARGOS [14]. Plus récemment, [11] ont montré que Prolog peut être utilisé pour représenter des théories scientifiques et permettre de nouvelles inférences (cf.table 4).

On voit bien là l'intérêt, sinon d'adopter des approches proposées par l'IA symbolique (lorsque cela est possible et pertinent), ou au moins de considérer des solutions hybrides, dites « neuro-symboliques ».

En ce sens, le langage Prolog, qui a été pensé dès sa conception comme un outil de l'intelligence artificielle symbolique, a toute sa place. On peut aussi noter que Prolog s'interface très facilement avec d'autres langages comme Python, actuellement très prisé par la communauté IA, et peut donc servir à la réalisation « d'IA neuro-symboliques ».

```
/* Si la dérivée de U par rapport à X est DU et dérivée
de V est DV, la dérivée de U + V est alors DU + DV, de
U−V est DU−DV, de U∗V est DU∗V+U∗DV, etc... */
d(U+V,X,DU+DV) :- d(U,X,DU), d(V,X,DV).
d(U-V,X,DU-DV) :- d(U,X,DU), d(V,X,DV).
d(U*V,X,DU*V+U*DV) :- d(U,X,DU), d(V,X,DV).
d(U/V,X,(DU*V-U*DV)/V^2) :- d(U,X,DU), d(V,X,DV).
```

TABLE 4. Comment formaliser les règles de dérivation de fonctions mathématiques pour, ensuite, demander à Prolog d'en déduire la dérivée d'une fonction.

Conclusion

Les principes mis en œuvre dans Prolog — unification, SLD Résolution (i.e. *Selective Linear Definite clause resolution* [8], logique du premier ordre, gestion éventuelle de contraintes sur des domaines — en font un langage original. Il est ainsi particulièrement bien adapté pour la résolution de problèmes qui peuvent s'exprimer sous forme de règles, de bases de données ou de faits.

En outre, il permet la modélisation d'une forme de raisonnement avec une économie d'efforts qui le rend séduisant. Il est simple à apprendre et à mettre en œuvre. Ses concepts et les instructions à maîtriser sont réduits ; en premier langage d'apprentissage, il présente de nombreux avantages, notamment du fait qu'il ne nécessite pas d'imaginer d'algorithme pour résoudre un problème. De même, l'apprentissage de la récursivité est « historiquement » intuitif en Prolog (voir par exemple [5]). Paradoxalement, il devient plus difficile à utiliser pour un programmeur formé aux boucles, aux structures typées et ayant l'habitude de concevoir un algorithme plutôt que de décrire le problème posé.

Prolog peut s'articuler aisément avec d'autres mécanismes d'IA, et en particulier les processus d'apprentissage de type connexionniste. Ainsi, Prolog peut générer facilement des contenus syntaxiquement exacts à partir de grammaires avec une économie d'énergie incomparable. En revanche, les aspects sémantiques ou qui relèvent de la perception (images, sons, etc) y sont plus complexes à représenter.

Cette combinaison contribue à la construction de solutions « intelligentes ». Cette approche hybride nous paraît représenter certains fonctionnements du raisonnement humain et elle ne s'oppose pas aux techniques « statistiques », elle les enrichit.

Les extensions de la programmation logique et ses dérivées ont démontré leur efficacité comme, par exemple, dans NukkAI avec la programmation logique probabiliste [9], dans BitCoinLog ou encore la vérification de protocole cryptographique [1].

Pour toutes ces raisons, il nous semble que les spécificités de Prolog (et de ses descendants comme les langages de programmation par contraintes, par exemple) esquissées ici sont un atout pour de nombreux problèmes nécessitant un cheminement explicable, pouvant être décrit sous forme de règles ou de relations entre entités, où le raisonnement vient compléter des processus d'apprentissage automatique, ou bien encore quand les données sont en volume réduit.

Références

[1] B. Blanchet et al. An efficient cryptographic protocol verifier based on prolog rules. In *csfw*, volume 1, pages 82–96, 2001.

[2] A. Colmerauer. Les bases de Prolog IV. *Le Manuel de Prolog IV, PrologIA*, 1996.

[3] A. Colmerauer, H. Kanoui, R. Pasero, and P. Roussel. Un système de communication en français, rapport préliminaire de fin de contrat IRIA. *Groupe Intelligence Artificielle, Faculté des Sciences de Luminy, Université Aix-Marseille II, France*, 1972.

[4] P. Deransart, A. Ed-Dbali, and L. Cervoni. *Prolog : the standard : reference manual*. Springer Science & Business Media, 1996.

[5] B. S. Elenbogen and M. R. O'Kennon. Teaching recursion using fractals in Prolog. In *Proceedings of the nineteenth SIGCSE technical symposium on Computer science education*, pages 263–266, 1988.

[6] L. Gouzènes. New Prolog Application Census 2022, 2022.

[7] R. Kowalski. Predicate logic as programming language. In *IFIP congress*, volume 74, pages 569–544, 1974.

[8] R. Kowalski and D. Kuehner. Linear resolution with selection function. *Artificial Intelligence*, 2(3-4) :227–260, 1971.

[9] J. Li, S. Thepaut, and V. Ventos. StarAI : Reducing incompleteness in the game of Bridge using PLP. *arXiv preprint arXiv :2001.08193*, 2020.

[10] D. Pastre. MUSCADET : An automatic theorem proving system using knowledge and metaknowledge in mathematics. *Artificial Intelligence*, 38(3) :257–318, 1989.

[11] J.-C. Rohner and H. Kjellerstrand. Using logic programming for theory representation and scientific inference. *New Ideas in Psychology*, 61 :100838, 2021.

[12] R. Schumi and J. Sun. ExAIS : Executable AI Semantics. *arXiv preprint arXiv :2202.09868*, 2022.

[13] R. Schwartz, J. Dodge, N. A. Smith, and O. Etzioni. Green AI. *Communications of the ACM*, 63(12) :54–63, 2020.

[14] J.-P. Spagnol. ARGOS, un demonstrateur de theoremes en geometrie. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, 8(1) :113–125, 2001.

Algol 60, 60 ans après

Thierry Dumont [1]

L'anniversaire d'Algol semble être un peu passé inaperçu. Pourtant il méritait bien qu'on s'intéresse à lui, ce grand-père de la plupart des langages de programmation procéduraux contemporains. L'auteur des ces lignes a récupéré cinq livres dans les rejets de la bibliothèque de mathématiques de son université : deux volumes de [15] [2] *(des programmes écrits dans les années soixante), et trois tomes d'*Actes de congrès de l'AFCAL(TI) [3] *(1960, 61 et 63) qui ensemble donnent une bonne idée de la pratique du calcul et de la programmation d'il y a 60 ans. Le point de vue ici est différent de celui de* [7]*, qui s'intéresse à la réalisation des premiers compilateurs : peut-on étudier, compiler et exécuter les programmes de* [15]*, apprendre et pratiquer le langage Algol 60 ? Les* Actes des congrès de l'AFCAL(TI) *permettent-ils d'évaluer l'adaptation du langage aux pratiques et aux besoins de l'époque de leur publication ? Enfin, comme* [15] *présentent des résultats de calculs, peut-on extraire des connaissances de ces résultats ?*

1. Institut Camille Jordan.
2. Produit des Recherches coopératives sur programme (RCP) 30 et 136 du CNRS.
3. Association française du calcul, puis Association française du calcul et de traitement de l'information.

DOI:10.48556/SIF.1024.20.65 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 65–76

Une présentation rapide d'Algol 60

Naissance

En octobre 1955, se tient à Darmstadt un symposium international d'où ressort la nécessité d'unifier les langages qui commencent à se développer dans le plus grand désordre [4] vers un modèle de langage indépendant des machines. Des meetings se succèdent : à l'automne 1957 [5], en avril 58 (meeting de l'ACM) et enfin en mai et juin 58 à Zurich. Ils aboutissent à la définition d'Algol 58. Malheureusement, Algol 58 est plein d'inconsistances : lors de réunions à Paris en novembre 59 et janvier 60, auxquelles participent entre autres J. Backus, P. Naur, H. Rutishauser et le français B. Vauquois, on procède à une reconception complète du langage qui devient *Algol 60*. Mais des inconsistances persistent et, après des meetings à Münich (1962) et à Delft (1963), sort sous l'égide de l'IFIP [6] un *Revised Algol Report*. Finalement, les difficultés d'implantation mènent l'IFIP à créer un *subset of* Algol 60. On trouvera une histoire partielle mais beaucoup plus détaillée dans P. Naur [13].

Le langage

Le texte de H. Rutishauser [16] est une remarquable description d'Algol 60 et aussi un magnifique cours de programmation, très pédagogique, accompagné d'exemples provenant tous du calcul scientifique (de l'époque). Il contient aussi la description du langage sous la forme de Backus-Naur (développée précisément pour Algol). Le langage décrit est celui du *subset of* Algol 60. On pourra aussi consulter le *Revised Report on Algorithmic Language Algol 60* [14].

La présentation ci-dessous suppose que le lecteur est un peu familiarisé avec les langages procéduraux actuels. On ne détaillera pas ce qui fait aujourd'hui partie d'une « culture générale » de la programmation, culture qui n'existait pas en 1960.

Types

Ils sont en nombre très limité (3) : `real, integer, boolean`. On remarque qu'il n'y a qu'une seule précision pour les flottants, et pas de type `complex`. On peut construire des tableaux de ces types primitifs avec des bornes entières quelconques :

```
real array a[-5:30];
integer array b,c[0:50,-11:30];
boolean array bl[0:20];
```

4. Organisé par la *Geselshaft für angewandte Mathematik und Mechanik* (GAMM).
5. Organisé par la GAMM et l'*Association for Computing Machinery* (ACM).
6. *International Federation for Information Processing*.

(ici `b` et `c` sont deux tableaux de mêmes bornes). Les limites sont incluses.

Algol 60 connaît les `label` (étiquettes) et les `switch` ; ces derniers sont très semblables aux `switch` du langage C. On verra plus loin que les uns et les autres peuvent être arguments de procédures.

Structures de contrôle

Elles sont au nombre de trois seulement !

(1) `if`. Voici deux exemples :

```
if x<0 then y:=1;
if x<0 then y:=1 else
begin
        z:=1;
        w:=3
end
```

(2) La boucle `for`. Elle n'opère pas forcément sur des entiers (à vos risques et périls !) :

```
for x:=10 step -0.1 until 0.2 do
        s:=s+exp(x);
```

La boucle `for` est peut-être la partie la plus étrange du langage. Ainsi :

```
for z:=10,11,9,8 do outreal(1,z);
```

va produire : 10 11 9 8 [7]. Mais, voici plus délicat :

```
integer k;
for k:= 10,k/3 while k>0 do
outinteger(1,k);
```

va produire : 10 3 1 ; ce qui correspond à la boucle `for(k=10;k>0;k=k/3)` en langage C.

Algol permet des constructions intéressantes :

```
real array a[1:n,1:n];
for i:=1 step 1 until n do
for j:=1 step 1 until n do
a[i,j]:= if j<i then 0 else 1;
```

(3) Enfin, Algol 60 possède l'infâme `goto`, mais on peut en limiter considérablement l'usage : par exemple l'utiliser pour sortir conditionnellement d'une boucle car il n'y a pas d'instruction *break*.

7. Notons que les entrées-sorties n'étaient pas normalisées en Algol 60. `outreal` n'est qu'une réalisation possible.

Blocs

Les blocs `begin...end` permettent d'intéressantes allocations dynamiques :

```
n:=10;
begin
        integer array y[if n=1 then 0 else 2:
        if n=10 then n else 2*n];
end;
```

`own` permet de déclarer des variables persistantes dans un bloc :

```
begin
        own integer array y[0:20];
end;
```

Procédures

Donnons l'exemple d'une procédure qui calcule un zéro approximatif d'une fonction strictement croissante par bissection en supposant que $f(a) < 0$ et $f(b) \geqslant 0$:

```
procedure bisect(eps,f,a,b);
real a,b,eps; real procedure f;
begin
        real x;
        for x:= (a+b)/2
                while b-a>eps do
                        if f(x)<0 then a:=x else b:=x;
end bisect;
```

Algol 60 permet d'emboîter les procédures (une procédure `B`, dans une procédure `A` n'est pas visible de l'extérieur de `A`). Par défaut, les paramètres sont passés par référence, mais on peut les passer par valeur en les déclarant `value`. Ainsi, dans l'exemple précèdent, on aurait pu écrire :

```
procedure bisect(eps,f,a,b);
value eps;
real a,b,eps; real procedure f;
```

Les paramètres effectifs correspondants à `a` et `b` dans le programme appelant sont modifiables mais pas celui correspondant à `eps`. Passer une étiquette comme paramètre permet, entre autres, de simuler un système d'exceptions, indisponible en Algol 60 :

```
integer procedure facto(n,exception);
value n;
integer n; label exception;
begin
        if n<0 then goto exception;
```

```
        if n=1 or n=0 then
                facto := 1
        else
        facto:= n*facto(n-1,exception);
end;
begin
integer n,k;
        for n:=12 step -1 until -1 do
                k:= facto(n,excep);
        goto fin;
        excep: outstring(1,"erreur.");
fin: end
```

C'est à peu près tout ! Avec les quelques lignes ci-dessus et le texte de Rutishauser [16], le lecteur doit pouvoir programmer en Algol 60. Mais, comment ne pas citer les mots de C.A.R. Hoare : « *Voici un langage très en avance de son temps, il n'a pas seulement été une amélioration de ses prédécesseurs mais aussi une amélioration de presque tous ses successeurs* ».

La réception d'Algol 60 — Les premiers compilateurs

L'intérêt pour Algol et la floraison des compilateurs

L'intérêt pour Algol peut se mesurer à la liste des pays où le journal « Algol Bulletin » était reçu ; parmi ceux-ci les pays d'Europe de l'Ouest et d'Amérique du Nord, la Tchécoslovaquie, l'Allemagne de l'Est, la Pologne, l'URSS, le Mexique et la Chine Populaire. Une liste incomplète des nombreuses implantations est donnée dans [21] ; la réalisation d'un compilateur par C. Pair et ses collègues est décrite dans [7] ; les *Actes de l'AFCALTI* (1963) mentionnent aussi un compilateur pour Bull Gamma 60. On trouve même la trace de la réalisation d'un compilateur à l'université de Pékin [6].

L'échec

Algol a échoué à devenir un langage sinon universel, du moins un langage répandu, pour finalement disparaître. Les raisons sont nombreuses. Notons déjà qu'IBM a sorti un brouillon du langage Fortran en 1954, un premier compilateur dès 1957 et que des compilateurs Fortran existeront sur IBM 709, 650, 1620 et 7090, ce qui, ajouté à la puissance d'IBM à l'époque, donnait un sérieux avantage à Fortran. P. Mounier-Kuhn donne dans [12] une description détaillée de « l'ascension et la chute » d'Algol en France, de l'enthousiasme des universitaires, avec le soutien d'institutions comme le CNRS, jusqu'au recul et à l'abandon dans les plans industriels, tout en montrant le rôle fondateur de l'intérêt pour Algol dans le développement des bases de l'informatique en France. Dans [13], P. Naur explique

« qu'il était sous-entendu que le domaine [d'application du langage] était celui qui sera plus tard appelé *traitement de l'analyse numérique* ou même *calcul scientifique*. À cette époque, les autres applications commençaient juste à émerger, ..., tandis que le *calcul scientifique* pouvait reposer sur une vieille tradition de méthodes de calcul et d'analyse mathématique... »

Cette assertion mérite d'être revisitée. Qu'était le calcul scientifique en 1960 ? Les *Actes des congrès de l'AFCAL* permettent d'y répondre : en 1960, on calcule... ce qui est accessible, avec les limitations dues à la faible puissance et à la taille mémoire très limitée des calculateurs : systèmes linéaires de petite taille, interpolation, résolution de systèmes d'équations différentielles simples (avec des méthodes explicites). L'attaque frontale de problèmes de physique (par la résolution des équations aux dérivées partielles) est à peine envisagée. Calculer des tables numériques n'est pas encore désuet : J.M. Laborde réalisera des tables de fonctions en 1963 [10] !

En ce sens là, Algol 60 est parfaitement adapté au calcul scientifique de l'époque. L'existence d'un seul type de flottants le montre aussi : les problèmes résolus sont plutôt bien posés, ne serait-ce que parce qu'ils sont de petite taille, et il n'y a pas de réel besoin de précision étendue.

Mais on peut aussi remarquer qu'en 1963, l'intérêt du calcul scientifique se déplace vers la résolution des équations aux dérivées partielles [8]. En 1964, la NASA lance le développement de son code Nastran [4] (calcul de structures par éléments finis) entièrement en Fortran IV. On assiste au basculement de l'analyse numérique et du calcul scientifique vers la résolution numérique des équations aux dérivées partielles. Le livre de Temam [19] intitulé « Analyse Numérique » est un manifeste de *l'École Lions* : l'analyse numérique, et donc le Cacul scientifique, ce sera à présent la résolution numérique des équations aux dérivées partielles. Mais l'implantation de la méthode des éléments finis nécessite de manipuler des structures de données relativement complexes (maillages, matrices creuses, etc.) pour lesquelles on ne voit pas vraiment d'avantages à adopter Algol 60 plutôt que Fortran. Tous les grands codes éléments finis entrepris dès lors et jusqu'aux années 90 seront d'ailleurs programmés en Fortran (IV ou 77), pour le meilleur ou pour le pire (plutôt pour le pire).

Les calculs en précision étendue apparaîtront vite comme nécessaires, dès que la taille des problèmes à résoudre pourra augmenter (le conditionnement des systèmes linéaires croît comme h^{-2} (h : pas de discrétisation, taille des éléments finis...) pour des problèmes elliptiques comme l'équation de Poisson : ce sont donc des problèmes assez mal conditionnés).

Les enfants d'Algol 60

On ne s'intéresse ici qu'aux enfants directs, nés avant 1970. Créé par Ershov, en URSS à Novosibirsk, *Alpha* ajoute à Algol 60 des manipulations de tableaux et de

8. Se référer à la table ronde de l'AFCALTI sur les EDP dans les actes de 1963.

matrices avec des « slices ». Conçu comme successeur d'Algol 60, *Algol 68* [11] est extrêmement rigoureux, mais très complexe (« *no implementations and no users* »). C'est un échec. Au moins un compilateur existait au début des années 70 : celui de *Royal Radar* en Grande Bretagne. Conçu par N. Wirth, qui n'a pas voulu suivre le mouvement Algol 68, *Algol W* [18] a été en vogue jusqu'au début des années 80, au moins dans l'enseignement. Il ajoute les types `long real, complex, long complex, bits`, « un ensemble ordonné de valeurs simples » avec `record`, la notion de référence à un `record` avec `reference`, ainsi qu'une bibliothèque standard.

Développé au *Norvegian Computing Center*, *Simula 67* est probablement le premier langage à classes, implantant des coroutines et nombre de concepts qui deviendront populaires bien plus tard.

Les machines Burroughs, à partir du B5000 en 1961, étaient des machines à architecture de pile et leur système d'exploitation était programmé avec des extensions d'Algol 60, successivement appelées *Espol et Newp*, ceci bien avant Multics (programmé en PL/1) et les différents Unix (programmés en C). Notons que des compilateurs sont disponibles sous Linux pour Simula 67 [2] et Algol 68 [1].

Faire revivre des programmes écrits en Algol 60

Pour rejouer les programmes écrits en Algol 60, il faut disposer d'un compilateur. Deux au moins sont disponibles : *GNU Marst* qui est peut être un peu en déshérence et *jff-algol*, réalisé et maintenu par Jan van Katwijk, disponible sur GitHub [20] ; c'est ce compilateur que nous avons utilisé.

On choisit ensuite un programme dans [15], par exemple la méthode de Bairstow qui permet de trouver les racines d'un polynôme $P(x)$ à coefficients réels sans utiliser d'arithmétique complexe (absente d'Algol 60). Pour cela, on cherche un diviseur quadratique de P, et on réitère le processus jusqu'à avoir trouvé toutes les racines de P. Plus précisément, B et C étant donnés, on a, en divisant P par (x^2+Bx+C) :

$$P(x) = (x^2+Bx+C)\ Q(x) + (Rx+S),$$

ce qui définit une application $\mathscr{F} : (B,C) \mapsto (R,S)$. On cherche alors, par la méthode de Newton, un couple (B,C) qui annule R et S. Le calcul des dérivées partielles de R et S par rapport à B et C est donné dans [8]. Pour implanter la méthode, il suffit de savoir diviser des polynômes par des trinômes, la méthode de Newton ne nécessitant que la résolution de systèmes linéaires de taille 2.

La figure 1 montre quelques extraits du code, programmé par Lagouanelle à Toulouse (les programmes de [15] ne sont pas anonymes). Surprise : c'est de l'Algol en Français, ce qui est possible, car Algol 60 fait une distinction entre une forme du langage pour les publications et des formes « pratiques », tenant compte aussi des contraintes matérielles [9]. Comment transcrire ce code ? un OCR s'avère insuffisant :

9. Des caractères n'étaient pas disponibles sur certaines machines comme ^ ou [ ].

il faut s'armer de patience et tout compléter, corriger vérifier. Le problème des mots clés en français n'est pas grave : on peut écrire un petit script pour les transcrire en anglais, mais Jan Van Katwijk a ajouté une option à son compilateur pour traiter l'« Algol en français ».

Une fois le code compilé et exécuté, il faut vérifier, dans la mesure du possible, qu'il est correct. Pour cela une programmation de la même méthode dans un autre langage (ici en Julia) permet de vérifier dans une certaine mesure la correction du programme, aux erreurs de troncature près. On peut aussi s'intéresser à des cas où les coefficients sont rationnels (ou même algébriques) et utiliser SageMath [5] qui permet de calculer dans une extension algébrique.

Le programme semble bien être juste !

Que dire de la qualité de la programmation ? On remarque qu'il y a quatorze `goto` dans le code originel, et qu'il n'y a pas de découpage en sous-procédures (et donc pas de procédures imbriquées). Clairement, ce n'est pas le mieux programmé de la collection, ce qui contraste avec les codes réalisés par exemple à Grenoble, qui ont

```
PROCEDURE BAIRSTOW(N,A,S0,P0,EPS1,EPS2,MAXITER,X,Y,DIVERG) ;
          VALEUR N,A ; ENTIER N,MAXITER ; REEL S0,P0,EPS1,EPS2;
          TABLEAU A,X,Y ; ETIQUETTE DIVERG ;
       COMMENTAIRE
        BAIRSTOW CALCULE LES RACINES DE P(X)=A(0)X**N+A(1)X**(N-1)+..A(N)
        POLYNOME DE DEGRE N.SES COEFFICIENTS,REELS,SONT RANGES DANS LE

      DEBUT TABLEAU B,C[-2:N] ;
            REEL R,T1,Y1,Y2,S,P,S1,P1,DELTA,DISCR,S2,P2,X1,X2 ;
            ENTIER I,K,NITER ;
             B[-2]:= B[-1]:=0 ;
             C[-2]:= C[-1]:=0 ;
             T1 := EPS1 ;
            SI N=0 ALORS ALLERA  FIN ;
            SI N=1 ALORS DEBUT B[0]:=A[0]; B[1]:=A[1];
                                ALLERA  DERZERO
                         FIN ;
       TESTO : I:=N+1 ;
               POUR I:=I-1 TANTQUE  A[I]=0 ET I > 1 FAIRE
               DEBUT X[I]:=0 ;Y[I]:=0 ;N:=N-1 ;
               FIN ;
       DEGTEST: SI N > 2 ALORS ALLERA  BOUCLE ;
                 S2:=-A[1]/A[0] ;
                 P2:= A[2]/A[0] ;
               ALLERA  ETIQ2 ;
               S:=S0 ; P:=P0 ;
       BOUCLE : EPS1 :=T1 ; NITER :=0 ;
               SI A[N-2]=0 ALORS DEBUT
               SI (ABS(S)+ABS(P)=0) ALORS DEBUT
               S:=0.1 ; P:=0.2 ; FIN FIN ;
         ITER : NITER :=NITER+1 ;
```

FIGURE 1. Méthode de Bairstow (extrait)

tous à peu près le même style, excellent, faisant probablement apparaître une culture locale commune de la programmation.

On peut aussi reprogrammer la méthode de Bairstow en Algol 60, en s'inspirant du code écrit en Julia. On remarque que le code Julia se transcrit facilement en Algol, avec *un seul* `goto`. La totalité de ces morceaux de code, ainsi que le code originel, sont disponibles sur GitHub [9].

Une expérience « d'archéologie numérique »

Dans les *Procédures*, il n'est pas fait mention des machines utilisées, qui diffèrent pourtant d'un centre à un autre. Les nombres flottants ne sont pas normalisés à l'époque, pas plus que la taille de leur mantisse. On peut tenter de deviner le type de machine utilisé en se servant des résultats fournis.

Ainsi, quelle était la machine utilisée à Grenoble avant 1967 (avant l'arrivée de l'IBM 360) ? Page 280 du tome 1 des *Procédures*, on trouve le code d'un ajustement polynomial aux moindres carrés, écrit par André Eberhard (auteur aussi d'une remarquable Transformée de Fourier rapide). L'auteur utilise une méthode naïve : ajuster les coefficients a_i de $p(x) = \sum_{i=0}^{m} a_i x^i$ en résolvant $\min_{\{a\}} \sum_{i=1}^{n} (p(x_i) - y_i)^2$. C'est une méthode très instable car la matrice du système linéaire à résoudre est un avatar de la matrice de Hilbert [10]. De plus, le système linéaire est résolu en utilisant la routine `gresolsyslin` (Tome 1 des *Procédures*, due à N. Gastinel). C'est une élimination de Gauss sans choix du pivot, ce qui est aussi source d'instabilité. Mais l'instabilité numérique peut être utile...

Quelle était donc la précision des flottants utilisés ?

Le problème résolu dans les *Procédures* consiste à ajuster un polynôme de degré 4 parmi 21 points (x_i, y_i) où les x_i sont équirépartis dans $[0,1]$ et $y_i = \exp(x_i)$. L'auteur trouve pour $a_i, i = 0,4$ les valeurs 0.10003376×10, 0.99835205, 0.50927734, 0.14135742 et $0.69702148 \times 10^{-1}$. On se sert alors de SageMath [5] qui, utilisant la bibliothèque Mpfr [3], permet de calculer avec des flottants de précision quelconque. On transcrit les programmes des *Procédures* dans SageMath obtenant ainsi un programme à peu près analogue au code Algol, mais paramétré par la précision des flottants (le nombre de bits de la mantisse). Le conditionnement [17] du système linéaire qu'on résout est de 689500, d'où on déduit qu'en calculant avec 350 bits de précision, on obtient une solution a^{ref} avec une erreur relative de l'ordre de 10^{-100} sur les coefficients a, donc pratiquement exacte. Définissant la distance à la solution « exacte » par $\mathrm{d(a)} = \max_{\mathrm{i}=0,4} |\mathrm{a_i} - \mathrm{a_i^{ref}}|$, on peut calculer cette distance en fonction de la précision du calcul. La distance de la solution calculée dans les *Procédures* à la solution « exacte » est de $1.48724517163778\ 10^{-3}$. En reportant cette distance sur

10. On trouve aussi dans [15] une implantation de la bonne méthode utilisant des polynômes orthogonaux.

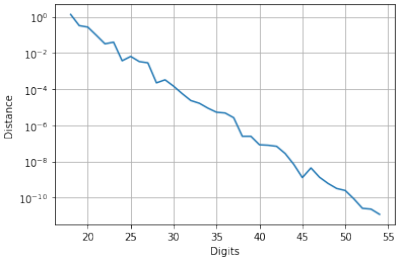


FIGURE 2. Distance / Précision.



FIGURE 3. Dr. Follamour : un IBM 7044.

le graphique de la figure 2, on trouve que le flottant de la machine utilisée avait une précision de 27 bits.

Quelles étaient les machines dotées d'un flottant avec une mantisse de 27 bits ? Il y avait des Dec PDP, quelques machines chez Univac et, chez IBM, la série 7000. Les PDP étaient des minis (des machines de laboratoire) et Univac était très peu répandu dans le monde de la recherche. Il reste IBM et on se souvient alors qu'il existait à cette époque un Centre de recherche IBM à Grenoble qui a beaucoup collaboré avec l'IMAG. La machine était donc un IBM, série 7000, plus précisément un IBM 7044 [11] (une machine célèbre, voir la figure 3).

Les IBM de la série 7000 avaient des mots de 36 bits et une mémoire pouvant aller jusqu'à 32 kmots. En simple précision, la mantisse avait une taille de 27 bits, mais il existait une double précision avec une mantisse de 54 bits [12]. En 1963, IBM a introduit des machines moins coûteuses, avec moins d'instructions dont faisait partie

11. Jean-François Maître (Pr. émérite, École centrale Lyon) a retrouvé un article de la presse locale l'annonçant.

12. Algol 60 avait un seul type de flottants, mais les compilateurs Fortran de l'époque avaient une option pour passer tous les flottants en « double ». Une option identique aurait certainement été possible pour Algol 60.

le 7044. Parmi les fonctionnalités sacrifiées, il y avait... la double précision. Pourquoi n'a-t-on pas acheté une machine munie de la double précision ? Peut-être qu'on n'en avait pas les moyens, car les unités flottantes étaient fort coûteuses à l'époque, ou peut être qu'on s'est dit qu'on n'en avait pas besoin vu les calculs qu'on projetait d'effectuer (voir plus haut !). Mais la vérité est peut être un mélange des deux !

Références

[1] Genie algol 68g. http://progopedia.com/implementation/algol68g.

[2] Gnu cim. https://www.gnu.org/software/cim/.

[3] The gnu mpfr library. https://www.mpfr.org/.

[4] Nastran. https://fr.wikipedia.org/wiki/Nastran.

[5] Sagemath. https://www.sagemath.org/.

[6] Chen kunqiu, prix xia peisu (china computer federation). https://www.ccf.org.cn/Computing_history/Updates/2021-04-28/727169.shtml, 2020.

[7] Marion Créhange, Pierre Lescanne, and Alain Quéré. L'informatique de claude pair. https://doi.org/10.48556/SIF.1024.18.91.

[8] Thierry Dumont. Bairstow's method. https://github.com/Thierry-Dumont/Vintage_Computing/blob/master/Bairstow/Doc/bairstow.pdf.

[9] Thierry Dumont. Vintage computing. https://github.com/Thierry-Dumont/Vintage_Computing, 2019.

[10] Jean Marcel Laborde. *Tables Numériques de Fonctions élémentaires ...* Dunod, 1963.

[11] C.H. Lindsey and S.G. Van der Meulen. Informal introduction to algol 68. http://www.softwarepreservation.org/projects/ALGOL/book/Lindsey_van_der_Meulen-IItA68-Revised.pdf, 1977.

[12] Pierre Mounier-Kuhn. Algol in france : From universal project to embedded culture. *IEEE Annals of the History of Computing*, 36(4) :6–25, 2014.

[13] P. Naur. The european side of the last phase of the development of algol 60- acm sigplan. https://dl.acm.org/doi/pdf/10.1145/960118.808370?download=true, 1978.

[14] Peter Naur (editeur). Revised report on algorithmic language algol 60. http://homepages.cs.ncl.ac.uk/cliff.jones/publications/OCRd/BBG63.pdf.

[15] etc Noël Gastinel, Préfacier. *Procédures Algol en analyse numérique*. Éditions du CNRS, 1967.

[16] Heinz Rutishauser. Description of algol 60. http://www.algol60.org/docs/Rutishauser_Description_of_ALGOL_60_1967.pdf, 1967.

[17] Michelle Schatzman. *Analyse numérique : cours et exercices pour la licence*. InterEditions, 1991.

[18] Richard L. Sites. Algol w, reference manual. http://www.algol60.org/docsW/AlgolW_STAN.pdf.

[19] Roger Temam. *Analyse numérique*. Presses universitaires de France, 1970.

[20] Jan van Katwijk. Algol 60 compiler. https://github.com/JvanKatwijk/algol-60-compiler.

[21] Wikipedia. Algol 60. https://en.wikipedia.org/wiki/ALGOL_60.

Catala : formaliser la loi grâce à un langage de programmation dédié

Denis Merigoux [1] et Liane Huttner [2]

Ce texte est adapté d'une présentation orale donnée au séminaire Codes Sources du 10 mars 2022.

Introduction : programmer la loi ?

Transformer les textes de loi en programmes informatiques est une tâche cruciale dans l'administration publique, nécessaire à l'application massive du droit pour produire les millions de décisions individuelles que constituent le prélèvement des impôts ou la distribution des aides sociales. C'est cette tâche de transformation à laquelle nous nous intéressons du point de vue des méthodes formelles. Prenons tout de suite un exemple pour cadrer notre sujet d'étude :

> **Code Général des Impôts, Article 197, I, 3, b, 3° (circa 2019)**
>
> Le taux de la réduction prévue au premier alinéa du présent b est de 20 %. Toutefois, pour [...], le taux de la réduction d'impôt est égal à 20 % multiplié par le rapport entre :
>
> — au numérateur, la différence entre 20 500 €, pour [...], ou 41 000 €, pour [...], et le montant des revenus mentionnés au troisième alinéa du présent b, etc. ;
>
> — au dénominateur, 2 000 €, pour [...], ou 4 000 €, pour [...].

1. Inria Paris, denis.merigoux@inria.fr.
2. Université Panthéon-Sorbonne, liane.huttner@univ-paris1.fr.

DOI:10.48556/SIF.1024.20.77 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 77–84

Le texte ci-dessus est constitué d'extraits du code général des impôts (CGI), qui regroupe tous les textes législatifs et réglementaires qui décrivent entre autres la formule de calcul de l'impôt sur le revenu à partir de la déclaration de revenus. On peut remarquer que cet article de loi est assez différent de l'idée reçue que l'on peut se faire de ce genre de texte. Plus précisément, cet article présente une étape d'une décision se faisant sans intervention humaine, sans ambiguïté (apparente), et traitant des données quantitatives. En résumé, cet extrait du CGI décrit une étape d'un algorithme. D'ailleurs, s'il fallait traduire le texte ci-dessus en pseudo-code, cela donnerait un programme pas si trivial que cela :

```
TAUX = si (non [...(1)...]) alors 20 % sinon
  20 % * (
    ((si [...(2),(3)...] alors 20 500 € sinon 41000 €)
      - REVENUS) /
    (si [...(4),(5)...] alors 2000 € sinon 4000 €))
```

Ainsi, la loi fait parfois office de spécification algorithmique. Attention, toutes les lois ne sont pas des algorithmes ! Il est évidemment hors de question de programmer des notions fortement dépendantes au contexte et nécessitant le jugement humain comme la justice ou l'égalité. Mais certains domaines juridiques sont particulièrement adaptés à une interprétation algorithmique, comme le calcul de l'impôt ou des prestations sociales. Concrètement, voici un petit tour d'horizon des programmes informatiques en production dans les administrations, et qui appliquent des algorithmes spécifiés dans la loi :

Nom	Langage	Taille (lignes)
Impôt sur le revenu	M et C	100 000
Impôt sur les sociétés	Java	10 000
Cotisations sociales	SQL	20 000
Prestations sociales	COBOL	6 900 000
Allocation chômage	Java	1 300 000
Retraites	?	?
Droits de succession	?	?

Outre la taille tout à fait conséquente des programmes, on peut également remarquer que ceux-ci sont souvent écrits dans des langages de programmation *legacy* comme COBOL, ou inhabituels (20 000 lignes de SQL !) ou encore spécifiques à l'administration comme le langage M de la DGFiP [9]. Enfin, certains programmes sont toujours secrets, malgré l'obligation qui est faite aux administrations de publier leur code source depuis la loi pour une République numérique de 2016.

Pour autant, même publiés, ces programmes informatiques ne sont pas forcément lisibles et compréhensibles. Voici par exemple un extrait (lignes 347 à 363) du code de la DGFiP qui a calculé l'impôt sur les revenus 2020 [3] :

3. Publié sur `https://gitlab.adullact.net/dgfip/ir-calcul`, fichier `chap-cmajo.m`.

```
RDSTARDIF_DEF = max(0,
  FLAG_RETARD * (FLAG_TRTARDIF * RDBASE_MAJO
    + FLAG_TRTARDIF_R * SUPRD[00] + FLAG_TRTARDIF_F *
      (positif(TARDIFEVT2) * positif(RDSTARDIF_A +
        CODCOR-CODCOR_A - RDBASE_MAJO) *
      (positif(FLAG_RECTIF) * min(SUPRD[2],RDBASE_MAJO)
        + (1 - positif(FLAG_RECTIF)) *
          min(max(0,RDBASE_MAJO-SUPRD[00] *
            (1 - positif(FLAG_RETARD0718))),
              RDSBASE_REF)) +
      (1 - positif(TARDIFEVT2) * positif(RDSTARDIF_A
        + CODCOR-CODCOR_A- RDBASE_MAJO)) *
      (positif(FLAG_RECTIF) * min(SUPRD[00],RDBASE_MAJO) +
        (1 - positif(FLAG_RECTIF)) * RDSBASE_REF)) +
    (1 - positif(FLAG_TRTARDIF + FLAG_TRTARDIF_R +
      FLAG_TRTARDIF_F + FLAG_TRMAJOP)) *
    (positif(FLAG_RECTIF) * SUPRD[00] +
      (1 - positif(FLAG_RECTIF)) *
        (RDSTARDIF_A +CODCOR-CODCOR_A))));
```

Comment retrouver quel article de loi cette expression est censée modéliser ? Mission impossible si l'on est pas dans la tête de celui ou celle qui l'a écrit...

La situation française n'est pas une exception ; les problèmes de ce genre sont même parfois pires ailleurs. Aux États-Unis par exemple, le calcul de l'impôt sur le revenu repose toujours sur un programme écrit dans l'assembleur spécifique à un ordinateur des années 1960 [13] ! Par contre, il existe en France un exemple connu d'échec industriel en partie lié à la difficulté de traduire la loi en code informatique. Il s'agit du système de paie Louvois de l'armée, renommé en SourceSolde. Selon la presse [10, 12], l'État a profité de l'informatisation du calcul de la paie pour supprimer 750 postes au ministère des armées. Cependant, avec plus de 174 primes et indemnités différentes, impossible de définir des critères simples de calcul. S'en sont suivis des bugs catastrophiques avec des soldats et leurs familles privés de solde pendant plusieurs mois. Selon le rapport d'une mission d'information de l'Assemblée Nationale [1], « *pour certains, les défaillances du calculateur lui-même constituent le problème majeur à la source des dysfonctionnements de l'écosystème Louvois. Le chef d'état-major des armées, devant la commission, a ainsi déclaré : j'en viens à penser que c'est le cœur du système Louvois qui constitue le fond du problème : vraisemblablement, il a été mal, ou insuffisamment, spécifié* ». Le problème de la spécification de ces logiciels devant appliquer des textes de loi ou des règlements est donc cruciale. Mais alors, comment vérifier que ces spécifications, et le logiciel lui-même, sont fidèles à la loi ?

Des programmes fidèles à la loi ?

Une première solution évidente est de valider la spécification et le logiciel à l'aide de tests, écrits à la main par des juristes spécialistes du droit en question. Cependant, plusieurs raisons structurelles rendent très coûteux le maintien d'une telle base de tests. D'abord, l'espace des entrées des programmes basés sur la loi souffre souvent

d'explosion combinatoire. En effet, avec plus de 1500 variables d'entrées dans la déclaration de revenus, il faudrait un nombre astronomique de tests pour couvrir toutes les possibilités (certainement plus que de foyers fiscaux en réalité !). Ensuite, les juristes n'ont pas l'habitude d'écrire des tests et il est difficile de leur inculquer la notion de couverture de code ; dans les études de droit, on apprend à appliquer la loi, pas à inventer à l'avance des situations insolites pour en tester les recoins sombres, puisque ces situations sont produites gratuitement par la réalité. Enfin, la loi change régulièrement ; tous les ans ou plus pour la loi fiscale par exemple. Il faut donc remettre à jour toute la base de tests très régulièrement.

Pour toutes ces raisons, la traduction de la loi en code nécessite une méthode de validation complémentaire à celle des tests qui est trop coûteuse pour atteindre un niveau maximum d'assurance sur le logiciel.

Mais quelle méthode utiliser ? Naïvement, on pourrait se dire que le plus simple serait que les spécifications et le logiciel soient directement écrits par les experts du domaine, c'est-à-dire les juristes. Mais comment rendre cela possible en pratique ? Il semble difficile de demander à des juristes d'écrire des logiciels de plusieurs dizaines ou centaines de milliers de lignes de code. Au lieu de cela, est-il possible pour les juristes de relire le code source de ces logiciels afin de détecter des erreurs d'interprétation de la loi ? La réponse à cette question est compliquée car plusieurs obstacles se placent sur la route. Premièrement, un juriste peut-il lire du code informatique ? Le droit fait appel au raisonnement logique et on ne voit pas pourquoi un juriste ne pourrait pas comprendre ce qu'est une conditionnelle, ou même une fonction. Empiriquement, nos discussions interdisciplinaires ont fait ressortir que le problème n'était pas là, car fondamentalement, le raisonnement logique basique qui sous-tend un programme informatique est tout à fait accessible aux juristes.

Par contre, un juriste peut-il lire la syntaxe d'un langage de programmation ? Cela dépend du langage de programmation, et la syntaxe est effectivement une barrière à l'entrée qui rebute nombre de non-informaticiens. Cependant cet obstacle peut être levé soit avec une formation, soit par la conception de syntaxes adaptées et proches du langage naturel [3].

Finalement, le plus sérieux obstacle à ce qu'un juriste puisse relire du code informatique concerne l'organisation du code. En effet, le juriste se trouve dans la loi grâce aux numéros d'articles, d'alinéas, à la structuration en différents titres et chapitres des codes législatifs et réglementaires. Or, le code informatique possède sa propre logique d'organisation par unité computationnelle qui est souvent radicalement différente de celle du texte de loi. La correspondance entre article de loi et bout de code informatique qui le traduit est anarchique : parfois, un article de loi est traduit par une incise à dix endroits différents dans le code ; parfois, c'est l'inverse et un bout de code modélise en même temps dix articles de loi. Ainsi, quand un juriste relit du code informatique, il est très souvent perdu car il ne retrouve plus la structure familière du texte de loi qu'il est censé appliquer. Pour remédier à ce problème, les

administrations ont la plupart du temps recours à la rédaction d'un document intermédiaire entre le texte de loi et le code informatique, parfois appelé *spécification*. Il s'agit d'un texte écrit en français naturel, décrivant l'effet computationnel d'un texte de loi, et organisé *comme un programme*. Sauf que, quand les juristes *analystes* chargés de rédiger cette spécification sont peu familiers avec la programmation, cette spécification n'est souvent ni assez détaillée ni assez bien organisée pour être traduite telle quelle en code informatique...

Mais pourquoi donc est-il si difficile de faire coïncider la structure du code informatique et celle du texte de loi ? Il se trouve qu'il existe une discordance profonde, et de nature logique. Sarah Lawsky [4], rappelant récemment des découvertes de la communauté de la programmation logique dans les années 1980 [5], met en évidence la dichotomie entre cas de base et exceptions dans les textes du droit fiscal américain. En effet, la plupart des textes de loi énoncent d'abord un principe général (cas de base), puis une multitude d'exceptions venant nuancer l'application de ce principe suivant les situations. En reprenant l'exemple de l'article 197 du CGI, on peut surligner en gras des conséquences introduites par deux justifications en italique, une pour le cas général et l'autre pour une exception :

> *Le taux de la réduction prévue au premier alinéa du présent b* **est de 20 %**. *Toutefois, pour les contribuables dont les revenus du foyer fiscal, au sens du 1° du IV de l'article 1417, excèdent 19 176 €, pour la première part de quotient familial des personnes célibataires, veuves ou divorcées, ou 38 352 €, pour les deux premières parts de quotient familial des personnes soumises à une imposition commune, ces seuils étant majorés le cas échéant dans les conditions prévues au même premier alinéa,* **le taux de la réduction d'impôt est égal à 20 % multiplié par le rapport entre : [...]**

Computationellement, cette structure en cas de base et exception est difficile à représenter car avant d'évaluer le cas de base, il faut vérifier que l'on ne déclenche aucune exception. En informatique, le cas de base se trouve à la fin de la conditionnelle dans le `else` alors qu'il se trouve énoncé en premier dans le texte de loi. De plus, des structures logiques usuelles du droit plus compliquées comme des exceptions d'exceptions deviennent compliquées à exprimer avec des conditionnelles dans le code informatique. Si l'on veut que la structure du code informatique suive celle du texte de loi, il est alors impératif d'utiliser un langage de programmation embarquant une fonctionnalité inspirée de la logique *par défaut*, une forme de logique non-monotone [11]. Or, presqu'aucun langage de programmation ne présente de construction permettant d'encoder naturellement une telle logique par défaut.

Catala, une méthodologie et un langage

Partant de ces observations et de l'état de l'art, nous proposons d'abord une méthodologie pour la production et la maintenance des programmes informatiques dérivant d'une spécification législative ou réglementaire. Le but de cette méthodologie est de garantir la correction du programme par rapport à la loi, ou plutôt par rapport à l'interprétation de la loi choisie par les auteurs du programme [6].

Le sujet central de la méthodologie que nous proposons est le cadre de production ou de modification du code. Plutôt que d'être fait isolément par un programmeur à partir d'un document tiers écrit par un juriste (cycle en V), nous préconisons d'utiliser deux techniques issues des méthodes agiles : la programmation en *binôme* et la programmation *littéraire*.

Dans le premier cas, l'objectif est de faire dialoguer un duo juriste-programmeur autour d'une même source dans laquelle les deux parties peuvent se retrouver. Concrètement, une séance de traduction de la loi en code se passe de la façon suivante : dans une salle, deux personnes sont réunies — un juriste et un programmeur. Le juriste a préparé la séance à l'avance en réunissant tous les textes de loi et règlements qui encadrent le dispositif cible ; par exemple le calcul d'une prestation sociale. Idéalement, le juriste a compilé ces textes en les copiant-collant dans un fichier texte que le programmeur va récupérer.

Ensuite, dans le cas de la programmation littéraire, le programmeur traduit linéairement, en suivant l'ordre du texte de loi, article par article. À chaque alinéa de chaque article, une discussion s'engage entre le programmeur et le juriste : qu'y a-t-il à calculer ici ? Est-ce que cet article décrit une étape d'un algorithme qui nous intéresse ? Si la réponse est non, le programmeur laisse un commentaire et passe à l'alinéa suivant. Si la réponse est oui, le programmeur écrit une première version du code informatique correspondant à cet alinéa, directement en dessous du texte de loi dans le même document. Une fois la première version écrite, voire pendant l'écriture, le juriste relit le code et intervient : qu'est-ce qu'on calcule ici ? Que veut dire cette instruction informatique ? Est-ce que le code gère cette interprétation de la loi ? Le programmeur répond et la discussion se poursuit : comment fait-on pour gérer ce cas-limite ? Faut-il rajouter un champ d'entrée pour prendre une décision dans le calcul à cet endroit là ? Quelles seront les conséquences de nos choix pour les utilisateurs ?

Toutes ces questions et réponses mènent à la prise de micro-décisions qui, dans leur ensemble, reflètent une certaine interprétation du droit qui ne dit généralement pas son nom [7]. Il est donc crucial, et c'est le dernier point de notre méthodologie, que le code source résultant du processus de traduction de la loi en code soit *open source* afin que tous ceux qui sont concernés par l'application de cette loi puissent aller constater les choix d'interprétation de l'administration, et si besoin les contester. En cela, nous ne faisons que recommander de préserver la pratique administrative

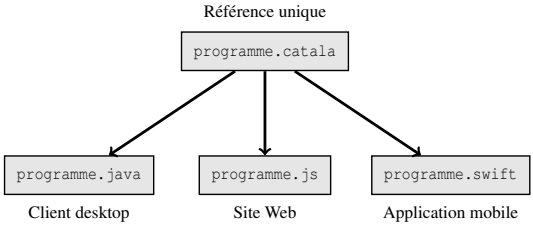
de publication des interprétations (par exemple au bulletin officiel des finances publiques).

Cette méthodologie est toutefois incomplète sans l'outil qui la rend possible. Il faut en effet disposer d'un langage de programmation qui permette de respecter la structure logique de la loi, qui soit compréhensible par le juriste, et qui dispose de suffisamment de capacité d'abstraction pour gérer des programmes de plusieurs centaines de milliers de lignes. C'est pour satisfaire ce cahier des charges que nous avons développé le nouveau langage de programmation spécialisé Catala [8, 2]. Nous ne décrirons pas les fonctionnalités du langage ni sa formalisation en détail ici, préférant vous renvoyer au papier technique sur le sujet [8]. Mais voici un petit exemple de traduction en Catala d'un article du code de la sécurité sociale décrivant une étape de calcul des allocations familiales :

> **Code de la sécurité sociale – Article L521-3** Toutefois, les personnes ayant un nombre déterminé d'enfants à charge bénéficient de ladite majoration pour chaque enfant à charge à partir de l'âge mentionné au premier alinéa.

```
champ d'application CalculAllocationsFamiliales :
  règle majorations_allocations_familiales.droits_ouverts
    de enfant
  sous condition
    (enfant dans ménage.enfants) et
    (nombre de ménage.enfants >= minimum_l521_3_alinéa_2) et
    (enfant.âge >= âge_limite_l521_3_alinéa_1 de enfant)
  conséquence rempli
```

Le code source Catala est ensuite compilé en passant par différentes représentations intermédiaires, qui vont jusqu'à des langages traditionnels en passant par du λ-calcul. L'idée est de disposer d'une référence unique pour toutes les implémentations de la loi dont auraient besoin les diverses applications informatiques d'un système d'information typique d'un grand ministère :



L'interopérabilité par livraison de code source directement dans le langage cible est plus contraignante qu'une interopérabilité par API, ou par *Foreign Function Interface* (FFI), mais c'est sans doute celle qui s'adapte le mieux à des environnements *legacy* où les technologies peuvent être très hétérogènes. De plus, la maîtrise de la chaîne de compilation permet de générer du code cible qui s'adapte parfaitement aux particularités de l'environnement où l'on veut le déployer.

Conclusion

Traduire la loi en code est une tâche ardue, méticuleuse mais néanmoins nécessaire et vitale pour le fonctionnement de nombre de services administratifs. Nous proposons une nouvelle méthodologie de production et de maintenance pour les programmes informatiques qui ont pour but d'implémenter une interprétation de la loi. Cette méthodologie s'appuie sur la programmation en binôme, la programmation littéraire, l'*open source* et un nouvel outil, le langage de programmation Catala [4].

Références

[1] Geneviève Gosselin-Fleury and Damien Meslot. Mission d'information sur la mise en œuvre et le suivi de la réorganisation du ministère de la Défense, 2013.

[2] Liane Huttner and Denis Merigoux. Catala : Moving Towards the Future of Legal Expert Systems. working paper or preprint, January 2022.

[3] Robert Kowalski. Logical english. *Proceedings of Logic and Practice of Programming (LPOP)*, 2020.

[4] Sarah B. Lawsky. Formalizing the Code. *Tax Law Review*, 70(377), 2017.

[5] L Thorne McCarty. Some arguments about legal arguments. In *Proceedings of the 6th international conference on artificial intelligence and law*, pages 215–224, 1997.

[6] Denis Merigoux. The Specification Problem of Legal Expert Systems. working paper or preprint, January 2022.

[7] Denis Merigoux, Marie Alauzen, and Lilya Slimani. Scrutinizing unnoticed programming choices in french housing benefits implementation. Submitted, 2022.

[8] Denis Merigoux, Nicolas Chataing, and Jonathan Protzenko. Catala : A programming language for the law. *Proc. ACM Program. Lang.*, 5(ICFP), August 2021.

[9] Denis Merigoux, Raphaël Monat, and Jonathan Protzenko. A modern compiler for the french tax code. https://doi.org/10.1145/3446804.3446850.

[10] Jacques Monin. Louvois, le logiciel qui a mis l'armée à terre, 2018.

[11] R. Reiter. Readings in nonmonotonic reasoning. chapter A Logic for Default Reasoning, pages 68–93. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.

[12] Philippe Reltien. Des milliers de soldats français endettés à cause de louvois, leur logiciel de paie. *France culture*.

[13] Tom Temin. IRS programming mystery continues. *Federal News Network*, 2020.

4. https://github.com/Catalalang/Catala.

Un langage pour enseigner la programmation, Scheme ou Python ?

Laurent Bloch [1]

Le texte qui suit est issu d'une séance du séminaire Codes Sources [5]. *Comme de juste il présentera des codes sources de différents programmes informatiques, mais aussi des codes sources de protéines et d'acides nucléiques, puisqu'aussi bien la cellule est un ordinateur (distribué) dont l'unité centrale (le ribosome) exécute les programmes de la vie, écrits selon le code génétique, dont nous donnerons aussi le source.*

Au cœur de la question

« *La programmation fonctionnelle est un assaut radical et élégant contre toute la pratique de l'écriture de programmes. Elle s'écarte totalement de la mentalité du programmeur qui* fait ci et puis après fait ça ». Vous devez recâbler votre cerveau d'une façon assez différente [2] ».

Faire ci et puis après faire cela

C'est la tendance spontanée du programmeur débutant, qui ne sait pas très bien par quel bout aborder le problème qu'il doit programmer. Elle est encouragée par le style de beaucoup de langages, surtout les langages de script en mode interprété. En général, cela finit par des séquences d'instructions pas forcément très cohérentes

1. https://laurentbloch.net.

2. Citation traduite de Simon Peyton-Jones, https://www.cs.cmu.edu/~popl-interviews/peytonjones.html

DOI:10.48556/SIF.1024.20.85

mais qui donnent des résutats parce qu'une heure avant, dans la boucle d'interaction de l'interpréteur, on aura donné aux variables impliquées dans le calcul des valeurs, on ne sait plus très bien lesquelles, comme le programme de gauche ci-après et, après une (courte) réflexion, comme celui de droite :

```
#!/usr/bin/env python3
x = 5
y = 0

if x != 0:
    y = x + x**2
    y = (y - 3) // x
    if y != 0:
        y = y % x
        print (x, y)
```

```
#!/usr/bin/env python3
# coding: utf-8

x = int(input("Entrez la valeur de x : "))
y = int(input("Entrez la valeur de y : "))
if x != 0:
    y = x + x**2
    y = (y - 3) // x
    if y != 0:
        y = y % x
        print (x, y)
```

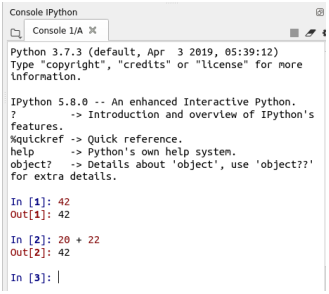
Les enseignants de l'option NSI des lycées n'en peuvent plus de ces programmes monolithiques à base d'`input` et de `print`. Pour un exposé raisonnable de l'enseignement de Python à des débutants, on pourra consulter le livre de Benjamin Wack et de ses coauteurs [6].

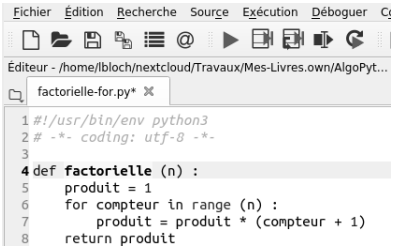
Pour construire un programme il faut...

Corrado Böhm et Giuseppe Jacopini [3] ont montré en 1966 que tout algorithme pouvait être programmé avec trois constructions : séquence d'instructions, alternative ou répétitive. Leur article est une des deux références de celui de Dijkstra [4], *Go to Statement Considered Harmful* (1968), l'autre est un article de Niklaus Wirth et d'Anthony Hoare sur le développement du langage Algol. L'art de la programmation réside pour une bonne part dans l'agencement harmonieux, agréable à lire et compréhensible de ces trois constructions, organisées pour constituer des fonctions.

Environnement de programmation

Ci-dessous quelques images d'un environnement de développement Python assez répandu, mais vous pouvez utiliser Emacs !





Biopython

Biopython est une bibliothèque de courts programmes Python (des *scripts*) qui permettent d'enchaîner de façon cohérente les programmes classiques de la bioinformatique (la plupart écrits en C) en les appliquant aux données du biologiste et aux banques de données du domaine. Biopython assure les conversions de format entre banques et fichiers et leur adaptation aux divers logiciels. C'est pour ce genre d'opérations que Python a été conçu ; il les effectue très bien et on se trompe en essayant de lui faire faire autre chose. Le programme qui suit, provenant du portail ExPASy de l'institut suisse de bioinformatique, correspond à la recherche de la séquence de la protéine identifiée par le numéro `O23729` (*Accession Number*). En général, les séquences biologiques sont identifiées sans ambiguïté par cet identifiant. Le lecteur trouvera le texte complet des programmes Biopython sur le site de l'auteur [1].

```
#!/usr/bin/env python3

from Bio import ExPASy
from Bio import SeqIO
with ExPASy.get_sprot_raw("O23729") as handle:
    seq_record = SeqIO.read(handle, "swiss")
    print(seq_record.id)
    print(seq_record.name)
    print(seq_record.description)
    print(repr(seq_record.seq))
    print("Length %i" % len(seq_record))
    print(seq_record.annotations["keywords"])
```

avec comme résultat d'exécution :

```
O23729
CHS3_BROFI
RecName: Full=Chalcone synthase 3; EC=2.3.1.74; AltName:
                        Full=Naringenin-chalcone synthase 3;
Seq('MAPAMEEIRQAQRAEGPAAVLAIGT...GAE', ProteinAlphabet())
Length 394
['Acyltransferase', 'Flavonoid biosynthesis', 'Transferase']
```

Biopython est un outil formidable, qui permet au biologiste d'enchaîner rapidement et facilement les analyses qu'il a l'habitude d'effectuer en routine. Mais aligner des lignes de Biopython n'est pas faire de la bioinformatique, c'est utiliser un outil sans vraiment comprendre ce qu'il fait, au risque d'être abusé par des résultats fallacieux. Disons que c'est un travail de technicien, et qu'un chercheur, à tout moment, peut être (doit être) son propre technicien.

Recherche de séquences

On peut essayer d'organiser un peu mieux ce travail en créant des fonctions qui pourront être réutilisées :

```
#!/usr/bin/env python3
def FetchSeqMulti(mydb, myrettype, myretmode, myid):
    from Bio import Entrez
    from Bio import SeqIO
    Entrez.email = "lb@laurentbloch.org"
    with Entrez.efetch(db=mydb, rettype=myrettype,\
                       retmode=myretmode, id=myid) as handle:
        for seq_record in SeqIO.parse(handle, "gb"):
            print("%s %s..." % (seq_record.id,\
                                seq_record.description[:50]))
            print("Sequence length %i, %i features, from: %s"
                  % (len(seq_record), len(seq_record.features),\
                     seq_record.annotations["source"]))

            FetchSeqMulti("nucleotide", "gb", "text",\
                          "6273291,6273290,6273289")
AF191665.1 Opuntia marenae rpl16 gene; chloroplast gene for c...
Sequence length 902, 3 features, from: chloroplast Grusonia marenae
AF191664.1 Opuntia clavata rpl16 gene; chloroplast gene for c...
Sequence length 899, 3 features, from: chloroplast Grusonia clavata
AF191663.1 Opuntia bradtiana rpl16 gene; chloroplast gene for...
Sequence length 899, 3 features, from: chloroplast Grusonia bradtiana
```

Code génétique

Dans la colonne de gauche du tableau ci-dessous, on trouve le nom des acides aminés du monde vivant et dans la colonne de droite, les motifs de trois acides ribonucléiques (ARN, un brin d'ARN est transcrit depuis un brin d'ADN, nucléotide pour nucléotide, en substituant un U (pour uracile) à un T (thymine)[3]) qui codent pour chacun (on note que ce code est redondant, ce qui procure des distractions aux bioinformaticiens). Dans la seconde colonne, le caractère qui code chaque acide aminé dans les banques de données (et dans les programmes qui les exploitent).

Alanine	A	Ala	GCU, GCC, GCA, GCG.
Arginine	R	Arg	CGU, CGC, CGA, CGG ; AGA, AGG.
Asparagine	N	Asn	AAU, AAC.
Acide aspartique	D	Asp	GAU, GAC.
Cystéine	C	Cys	UGU, UGC.
Glutamine	Q	Gln	CAA, CAG.
Acide glutamique	E	Glu	GAA, GAG.
Glycine	G	Gly	GGU, GGC, GGA, GGG.
Histidine	H	His	CAU, CAC.
Isoleucine	I	Ile	AUU, AUC, AUA.
Leucine	L	Leu	UUA, UUG ; CUU, CUC, CUA, CUG.
Lysine	K	Lys	AAA, AAG.
Méthionine	M	Met	AUG.
Phénylalanine	F	Phe	UUU, UUC.
Proline	P	Pro	CCU, CCC, CCA, CCG.
Pyrrolysine	O	Pyl	UAG, avant élément PYLIS.
Sélénocystéine	U	Sec	UGA, avec séquence SECIS.
Sérine	S	Ser	UCU, UCC, UCA, UCG ; AGU, AGC.
Thréonine	T	Thr	ACU, ACC, ACA, ACG.
Tryptophane	W	Trp	UGG. (UGA)
Tyrosine	Y	Tyr	UAU, UAC.
Valine	V	Val	GUU, GUC, GUA, GUG.
Initiation			AUG. (UUG, CUG)
Terminaison	*		UAG, UAA ; UGA.

3. https://fr.wikipedia.org/wiki/Transcription_(biologie)

Une protéine

Voici une protéine, une vraie, pas les choses un peu visqueuses que vous pouvez acheter dans les salles de gymnastique. Elle est issue d'une grenouille griffue d'Afrique de l'Ouest. Chaque ligne du code source ci-dessous est identifiée par un code de deux caractères en début de ligne. Le texte proprement dit de la séquence est à la fin, avec un code ligne vide. Le code ligne `//` marque la fin de la séquence. Le code ligne `ID` est un identifiant, peu utile parce qu'il peut changer au fil du temps. Le code ligne `AC` indique une ligne qui contient un ou plusieurs *Accession Numbers*, immuables, et de ce fait le premier est généralement utilisé pour identifier la séquence.

```
ID   1433B_XENTR             Reviewed;         244 AA.
AC   Q5XGC8; Q28HK2;
DT   22-NOV-2005, integrated into UniProtKB/Swiss-Prot.
DT   23-NOV-2004, sequence version 1.
DT   28-NOV-2006, entry version 18.
DE   14-3-3 protein beta/alpha.
GN   Name=ywhab;
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus; Silurana.
OX   NCBI_TaxID=8364;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RG   Sanger Xenopus tropicalis EST/cDNA project;
RL   Submitted (MAR-2006) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC   TISSUE=Embryo;
RG   NIH - Xenopus Gene Collection (XGC) project;
RL   Submitted (OCT-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Adapter protein implicated in the regulation of a large
CC       spectrum of both general and specialized signaling pathway. Binds
CC       to a large number of partners, usually by recognition of a
CC       phosphoserine or phosphothreonine motif. Binding generally results
CC       in the modulation of the activity of the binding partner (By
CC       similarity).
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasm (By similarity).
CC   -!- SIMILARITY: Belongs to the 14-3-3 family.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; CR760847; CAJ82973.1; -; mRNA.
DR   EMBL; BC084514; AAH84514.1; -; mRNA.
DR   UniGene; Str.8742; -.
DR   SMR; Q5XGC8; 1-231.
DR   Ensembl; ENSXETG00000022830; Xenopus tropicalis.
DR   InterPro; IPR000308; 14-3-3.
DR   Gene3D; G3DSA:1.20.190.20; 14-3-3; 1.
DR   PANTHER; PTHR18860; 14-3-3; 1.
DR   Pfam; PF00244; 14-3-3; 1.
DR   PRINTS; PR00305; 1433ZETA.
DR   ProDom; PD000600; 14-3-3; 1.
DR   SMART; SM00101; 14_3_3; 1.
DR   PROSITE; PS00796; 1433_1; 1.
DR   PROSITE; PS00797; 1433_2; 1.
KW   Acetylation.
FT   CHAIN         1    244       14-3-3 protein beta/alpha.
FT                                /FTId=PRO_0000058600.
FT   MOD_RES       1      1       N-acetylmethionine (By similarity).
SQ   SEQUENCE   244 AA;  27721 MW;  FF766793EA1CA9E5 CRC64;
     MDKSELVQKA KLSEQAERYD DMAASMKAVT ELGAELSNEE RNLLSVAYKN VVGARRSSWR
     VISSIEQKTE GNDKRQQMAR EYREKVETEL QDICKDVLGL LDKYLVPNAT PPESKVFYLK
     MKGDYYRYLS EVASGDSKQE TVTCSQQAYQ EAFEISKSEM QPTHPIRLGL ALNFSVFYYE
     ILNSPEKACS LAKSAFDEAI AELDTLNEES YKDSTLIMQL LRDNLTLWTS ENQGEEADNA
     EADN
```

Alignement de séquences

Les biologistes sont intéressés par les similitudes entre séquences, qui peuvent les mettre sur la piste de similitudes fonctionnelles, ou de parentés génétiques entre organismes. Le calcul consiste, pour le dire de façon très simplifiée, à affecter des coûts aux différences (mutations) et aux décalages *(gaps)* et à chercher l'alignement qui minimise la somme de ces coûts.

Voici un exemple jouet d'alignement de deux séquences nucléiques (ADN), avec le code source à gauche et l'éxécution à droite :

```
from Bio import Align

aligner = Align.PairwiseAligner()
seq1 = "GAACT"
seq2 = "GAT"
score = aligner.score(seq1, seq2)
alignments = aligner.align(seq1, seq2)
for alignment in alignments:
    print(alignment)
```

```
GAACT
||--|
GA--T
<BLANKLINE>
GAACT
|-|-|
G-A-T
<BLANKLINE>
```

Les deux solutions proposées comportent chacune deux vides (*gaps*) et trois concordances (*matches*), soit avec les paramètres par défaut de la méthode `PairwiseAligner()` un score égal à 3 (moins on additionne de coûts plus le score augmente).

BLAST bien sûr...

Le programme Blast, dont la première version a été publiée en 1990, est toujours le logiciel phare de la biologie moléculaire. Il peut être vu comme une extension et une généralisation des méthodes évoquées ci-dessus. Il est bien sûr accessible dans l'environnement Biopython.

```
#!/usr/bin/env python3
import sys
from Bio.Blast import NCBIWWW
from Bio.Blast import NCBIXML
def Blast_sonde(blastversion, collection, gi):
    result_handle = NCBIWWW.qblast(blastversion, \
                                   collection, gi)
    blast_record = NCBIXML.read(result_handle)
    E_VALUE_THRESH = 0.04
    for alignment in blast_record.alignments:
        for hsp in alignment.hsps:
            if hsp.expect < E_VALUE_THRESH:
                print("****Alignment****")
                print("sequence:", alignment.title)
                print("length:", alignment.length)
                print("e value:", hsp.expect)
                print(hsp.query[0:75] + "...")
                print(hsp.match[0:75] + "...")
                print(hsp.sbjct[0:75] + "...")

blastversion = sys.argv[1]
collection = sys.argv[2]
gi = sys.argv[3]
Blast_sonde(blastversion, collection, gi)
```

Si on connaît l'identifiant GenBank, avec la ligne de commande :

```
./BLAST-biopython.py blastn nt 8332116
```

BLAST évalue la pertinence statistique du score par une analyse de la distribution des scores d'alignement entre la séquence-test et l'ensemble des séquences cibles, et calcule la probabilité et l'espérance mathématique de trouver un alignement donnant un score donné parmi les cibles, uniquement du fait du hasard. L'espérance mathématique est notée *e-value*.

```
****Alignment****
sequence: gi|1219041180|ref|XM_021875076.1| PREDICTED: Chenopodium quinoa
cold-regulated 413 plasma membrane protein 2-like (LOC110697660), mRNA
length: 1173
e value: 1.63199e-117
ACAGAAAATGGGGAGAGAAATGAAGTACTTGGCCATGAAAACTGATCAATTGGC ...
|| ||||||||| |||| | |||| ||  |||| |||| | |||| ||| | || ...
ACCGAAAATGGGCAGAGGAGTGAATTATATGGCAATGACACCTGAGCAACTAGC ...
****Alignment****
sequence: gi|1226796956|ref|XM_021992092.1| PREDICTED: Spinacia oleracea
cold-regulated 413 plasma membrane protein 2-like (LOC110787470), mRNA
length: 672
e value: 1.0299e-113
AAAATGGGGAGAGAAATGAAGTACTTGGCCATGAAAACTGATCAATTGGCCGT ...
|||||||| |||  |||| | || ||||| |||||||| || ||||| ||||  ...
AAAATGGGTAGACGAATGGATTATTTGGCGATGAAAACCGAGCAATTAGCCGC ...
****Alignment****
sequence: gi|731339628|ref|XM_010682658.1| PREDICTED: Beta vulgaris subsp.
vulgaris cold-regulated 413 plasma membrane protein 2 (LOC104895996), mRNA
length: 847
e value: 2.76359e-108
TTGGCCATGAAAACTGATCAATTGGCCGTGGCTAATATGATCGATTCCGATAT ...
|||||||||||||||| ||| ||||  |||||||| |||| ||||  ||||| ...
TTGGCCATGAAAACTGAGCAAATGGCGTTGGCTAATTTGATAGATTATGATAT ...
****Alignment****
sequence: gi|1389679838|ref|XM_016034586.2| PREDICTED: Ziziphus jujuba
cold-regulated 413 plasma membrane protein 2-like (LOC107424728), mRNA
length: 946
e value: 1.43158e-105
AAAATGGGGAGAGAAATGAAGTACTTGGCCATGAAAACTGATCAATTGGCCGT ...
||||||||||    ||| |||  ||||| |||| |||||||| |   |||   ...
AAAATGGGGAGG---ATGGAGTTTTTGGCTATGAGAACTGATCCA---GCCAC ...
```

Apprend-on ainsi à programmer ?

L'expérience accumulée par les chercheurs dans le domaine des interfaces entre l'être humain et la machine (IHM) peut nous aider à comprendre les démarches d'apprentissage de la programmation. Ainsi, ils ont remarqué que la difficulté ne résidait en général pas dans l'écriture d'une instruction individuelle. Ce résultat est confirmé par l'expérience des tableurs, qui ont permis à des milliers de comptables de programmer des systèmes de gestion assez complexes en écrivant des formules de calcul dans des cases, ce qui en soi est un résultat remarquable. Ce constat peut être étendu à toutes sortes de *Frameworks*, *Content Management Systems* ou *Cascading Style Sheets*, qui permettent à des professionnels peu pourvus en compétences informatiques de réaliser les applications qu'ils souhaitent en insérant ou en modifiant une instruction ponctuelle dans un cadre plus vaste qui leur est fourni tout préparé.

Ce qui est vraiment difficile, malgré l'enseignement du *Discours de la méthode*, c'est de décomposer un problème en sous-problèmes, ce qui explique la propension du débutant à « faire ci et puis après faire ça », parce que la démarche itérative semble

contourner cette difficulté. Comme on le sait, ce contournement ne tarde jamais à se retourner contre son auteur, sous la forme du célèbre « plat de spaghettis » auquel va bientôt ressembler son programme. Ceci nous semble une bonne raison d'encourager le style de programmation fonctionnel, qui encourage et facilite cette décomposition des problèmes en sous-problèmes, résolus par des sous-programmes. Le concept de sous-programme nous semble ici un pivot de l'art de la programmation.

Le style fonctionnel est possible avec la plupart des langages, mais certains l'encouragent plus que d'autres.

Tri Rapide en Python

Le programme Python ci-dessous a l'avantage d'être lisible, ce qui en fait un bon pseudo-code. Ses performances sont pour le moins médiocres, et le rôle sémantique du caractère de tabulation pour dénoter la structure du code, s'il est acceptable pour un script bref avec un seul niveau d'indentation, est rédhibitoire pour de vrais programmes, des fonctions de plus de dix lignes, du texte que l'on voudra déplacer dans le corps du programme, etc.

```
from partition import partition

def tri_rapide (T):
    imin = 0
    imax = len(T) - 1
    tri_part(T, imin, imax)
    print (T)

def tri_part(T, imin, imax):
    if imin < imax:
        q = partition (T, imin, imax)
        tri_part(T, imin, q)
        tri_part(T, q + 1, imax)
```

```
from echange import echange

def partition (T, imin, imax):
    x = T[imin]
    while True:
        while T[imax] > x:
            imax -= 1
        while T[imin] < x:
            imin += 1
        if imin < imax:
            echange (T, imin, imax)
            imax -= 1a
            imin += 1
        else:
            return imax
```

```
def echange(V, i, j):
    tmp = V[i]
    V[i] = V[j]
    V[j] = tmp
```

Le lecteur trouvera le texte complet des programmes, y compris les fonctions auxiliaires pour mesurer les temps de calcul, sur le site de l'auteur [2].

Tri Rapide en Scheme

Le même tri est programmé en langage Scheme et présenté ci-après. Après avoir constaté l'allergie des biologistes aux manuels de programmation dont les exercices s'attaquent à la résolution numérique d'équations différentielles, l'auteur a rédigé son propre manuel, avec des exemples d'algorithmes bioinformatiques [**?**].

```
(define (tri-rapide v imin imax)
  (if (<fx imin imax)
      (let ((q (partition v imin imax)))
        (tri-rapide v imin q)
        (tri-rapide v (+fx q 1) imax)))
  v)

(define (swap v i j)
  (let ((x (vector-ref v i)))
    (vector-set! v i (vector-ref v j))
    (vector-set! v j x)))
```

```
(define (partition v imin imax)
  (let ((x (vector-ref v imin)))
    (let loop ()
      (let loop1 ()
        (if (>fx (vector-ref v imax) x)
            (begin
              (set! imax (-fx imax 1))
              (loop1))))
      (let loop2 ()
        (if (<fx (vector-ref v imin) x)
            (begin
              (set! imin (+fx imin 1))
              (loop2))))
      (if (<fx imin imax)
          (begin
            (swap v imin imax)
            (set! imin (+fx imin 1))
            (set! imax (-fx imax 1))
            (loop))
          imax))))
```

Tri rapide en C

Toujours le même programme mais dans le langage C ci-après.

```
void quicksort(int tableau[], int premier, int dernier){
  int i, j, pivot, temp;
  if(premier < dernier){
    pivot = premier;
    i = premier;
    j = dernier ;
    while(i < j){
      while(tableau[i] <= tableau[pivot] && i < dernier)
    i++;
      while(tableau[j] > tableau[pivot])
    j--;
      if(i < j){
    temp = tableau[i];
    tableau[i] = tableau[j];
    tableau[j] = temp;
      }
    }
    temp = tableau[pivot];
    tableau[pivot] = tableau[j];
    tableau[j] = temp;
    quicksort(tableau, premier, j-1);
    quicksort(tableau, j+1, dernier);
  }
}
```

Temps d'exécution du tri rapide

En considérant 200×10^6 valeurs prises au hasard entre 1 et 10^8, le tableau ci-après donne les temps d'exécution du tri rapide, respectivement en C compilé avec `clang` et l'option d'optimisation `-O3`, et avec le compilateur Scheme Bigloo, option d'optimisation `-O5`. Compte tenu du fait que Bigloo vient avec un environnement d'exécution *runtime* et notamment un glaneur de cellules (*garbage collector*, GC), on peut considérer ce résultat comme satisfaisant. Python et les autres compilateurs Scheme échouent à effectuer ce calcul.

C		Bigloo	
real	65,3 s	real	109,0 s

Choix d'un langage d'initiation à la programmation

Nous identifions ici plusieurs critères :

— la facilité d'apprentissage est la première qualité d'un langage choisi pour l'enseignement à des débutants ;
— une syntaxe sobre, régulière et expressive contribue à la facilité recherchée ;
— la lisibilité : les langages qui expriment leurs constructions par des mots sont plus faciles à apprendre que ceux qui les expriment par des signes cabalistiques, ne serait-ce que parce que les mots sont plus faciles à mémoriser ;
— le niveau d'abstraction du langage doit se placer au « juste milieu » :
  — ne pas exiger d'emblée la compréhension du fonctionnement de l'OS et du processeur ;
  — mais permettre la programmation d'opérations élémentaires telles que tri ou fonctions de condensation.
— le langage choisi doit encourager la structuration en sous-programmes, grâce à des constructions faciles à comprendre ;
— le langage doit fournir une gestion automatique de la mémoire (GC) ; il y a bien assez de choses à apprendre et à comprendre avant la gestion explicite de la mémoire ;
— le langage doit pouvoir être compilé : seul un programme exécutable constitué matériellement dans un fichier peut communiquer l'intuition d'un artefact qui calcule ;
— les environnements intégrés de programmation munis d'interfaces graphiques actionnées à la souris nous semblent des obstacles à la compréhension plutôt qu'autre chose. Outre qu'ils sont intrinsèquement compliqués et difficiles à maîtriser, ils dissimulent des choses qu'il faut au contraire voir.

Apports techniques du style fonctionnel

Outre les avantages formels, évoqués ci-dessus, que procure le style fonctionnel, il apporte des avantages techniques. Depuis quelques années les performances unitaires des processeurs progressent de moins en moins vite, et leur miniaturisation se heurtera de toute façon à des limites physiques.

L'amélioration des performances des ordinateurs et autres artefacts programmables passe désormais par la multiplication des organes de calcul, que ce soit par les processeurs multi-cœurs (en fait des multi-processeurs) ou par le recours aux processeurs graphiques (*Graphics Processing Unit*, GPU), qui sont en fait des unités de calcul vectoriel, capables d'appliquer simultanément une instruction unique à de multiples données rangées dans un vecteur.

L'informatique en nuages (*cloud computing*, infonuagique), en permettant de répartir un calcul sur un nombre indéterminé d'ordinateurs dont on ne sait pas où ils sont, suggère une évolution des architectures de calcul dans cette même direction.

Ces technologies stimulent le recours au calcul parallèle ; garantir qu'une tâche ne modifiera pas l'état du système pris comme hypothèse par une autre tâche est nécessaire au déterminisme du calcul ; le style fonctionnel limite le recours à l'affectation et par là facilite la satisfaction de cette condition.

Le marché florissant de la lambda-expression

C'est l'aboutissement (provisoire) des évolutions mentionnées à la section précédente, la lambda-expression, qui jusqu'à très récemment appartenait au vocabulaire exclusif des informaticiens théoriciens, devient une denrée commerciale de grande consommation.

Remerciements

Pour les méthodes bioinformatiques, je dois tout à William Saurin. Manuel Serrano m'a aidé pour les comparaisons de programmes dans des langages différents. Ils ne sont bien sûr en rien responsables des erreurs ou imprécisions de ce texte et de ces codes sources.

Références

[1] Laurent Bloch. Python et biopython. 2019.

[2] Laurent Bloch. Python, scheme, c. 2019.

[3] Corrado Böhm and Giuseppe Jacopini. Flow diagrams, Turing machines and languages with only two formation rules. *Communications of the ACM (CACM)*, 9(5), mai 1966.

[4] Edsger Wybe Dijkstra. Letters to the editor : Go to statement considered harmful. *Commun. ACM*, 11(3) :147–148, March 1968.

[5] Raphaël Fournier-S'niehotta, Baptiste Mélès, and Lionel Tabourier. Séminaire codes sources.

[6] Benjamin Wack, Sylvain Conchon, Judicaël Courant, Marc de Falco, Gilles Dowek, Jean-Christophe Fillâtre, and Stéphane Gonnord. *Informatique pour tous en classes préparatoires aux grandes écoles - Manuel d'algorithmique et programmation structurée avec Python*. Eyrolles, Paris, 2013.

(R)évolutions de l'informatique dans le sup : orientation et pratiques pédagogiques !

Bruno Mermet [1]

Le 25 mai 2022 avait lieu la journée Enseignement [2] de la Société informatique de France. Après deux années passées à nous rencontrer principalement via écran interposé, pouvoir se retrouver (ou faire connaissance) était un événement attendu de tous. Certes, la date retenue, juste avant le week-end de l'Ascension, peu après les « épreuves anticipées reportées » du baccalauréat, n'a pas permis à toutes les personnes intéressées de pouvoir venir, mais la cinquantaine de participants était ravie d'être là.

Après l'édition précédente qui avait abordé la question des débouchés en termes d'études pour les élèves ayant suivi la spécialité NSI, cette nouvelle journée était l'occasion de faire un bilan après la première année de ces nouveaux bacheliers dans le supérieur.

La journée fut ainsi divisée en trois temps. Une première partie fut la présentation des choix d'orientation des nouveaux bacheliers ayant suivi la spécialité NSI par Jean-Marie Chesneaux, inspecteur général à l'inspection générale de l'éducation, du sport et de la recherche (IGESR) en charge de NSI. Dans un second temps, les structures d'accueil de ces nouveaux étudiants furent présentées par des collègues issus de diverses composantes de formations. Enfin, l'après-midi fut consacrée à la

1. Maître de conférences.

2. https://www.societe-informatique-de-france.fr/les-journees-sif/les-journees-pedagogiques/journee-enseignement-2022/.

DOI:10.48556/SIF.1024.20.97 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 97–105

présentation de démarches pédagogiques et outils utilisés dans l'enseignement de l'informatique du primaire au supérieur.

L'informatique au lycée et en prépa, concours du CAPES NSI et de l'agrégation informatique [3]

Jean-Marie Chesneaux commence par replacer la spécialité Numérique et sciences informatiques (NSI) dans un contexte plus large. En effet, depuis la dernière réforme des collèges, l'informatique est au programme du cycle 4, répartie entre les programmes de mathématiques et de technologie, ce qui doit permettre de développer le début d'une pensée informatique. Ensuite, en seconde générale, le programme de SNT doit permettre d'aborder les quatre piliers de l'informatique que sont l'information, les algorithmes, les programmes et les machines, autour de sept thématiques, telles que la localisation ou la photographie numérique. Ce n'est qu'après cette initiation à l'informatique du tronc commun que les élèves intéressés peuvent approfondir leur connaissance de l'informatique en un ou deux ans grâce à la spécialité NSI, représentant environ 250 heures de formation pour les élèves suivant la spécialité sur les deux ans, avec une part importante laissée à la réalisation de projets (un quart du volume horaire).

NSI

Des disparités régionales Le ministère a laissé les académies et les lycées apprivoiser la réforme, et notamment la mise en œuvre de la spécialité NSI, à leur rythme. Aussi, il existe de fortes disparités régionales. Ainsi, suite à des remontées statiques provenant de 27 académies, il apparaît que la proportion de lycées publics proposant NSI varie de 23 % à 90 % selon les régions, avec une moyenne à 61 %. La proportion est un peu plus faible dans les établissements privés sous contrat (41 % en moyenne sur 25 académies). De fait, le pourcentage d'élèves suivant la spécialité en terminale varie également de 4,6 % à 11,6 %, avec une moyenne à 7,8 %.

Les disparités se retrouvent également parmi les enseignants de la spécialité, tant sur le nombre (entre 8 et 235), qui peut s'expliquer par les tailles très variables des académies, que sur le pourcentage de femmes (entre 0 % et 25 %, avec une moyenne à 13 %). Concernant la formation, une grande majorité des enseignants (80 %) est titulaire du DIU Enseigner l'informatique au lycée, mais on trouve des académies avec un taux bien plus faible (le minimum étant à 40 %).

3. Exposé de Jean-Marie Chesneaux : `https://www.societe-informatique-de-france.fr/wp-content/uploads/2022/07/SIF-2022-05-25.pdf`.

Les flux entrants En seconde générale à la rentrée 2021, il y avait 560 000 élèves, répartis sur 18 600 classes. Ainsi, cette année, ce sont 28 000 heures de SNT qui ont été dispensées chaque semaine.

Les effectifs d'élèves qui suivent NSI en première et en terminale depuis l'ouverture de la spécialité ont été les suivants :

	2019 – 2020	2020 – 2021	2021 – 2022
Première	31502	34797 (+10 %)	37871 (+9 %)
Terminale		13907	16184 (+16 %)

TABLE 1. Effectifs de la spécialité NSI.

Les effectifs sont certes en hausse constante, mais restent relativement faibles. Cependant, Jean-Marie Chesneaux souligne que si ce rythme de croissance se maintenait, on obtiendrait 80 000 élèves en 2028, ce qui permettrait d'être en phase avec les besoins de la société française.

La part de filles suivant la matière reste par contre dramatiquement faible : elle est déjà faible à l'entrée (18 % en première) et s'aggrave en terminale (14 %), les filles abandonnant plus la spécialité que les garçons entre la première et la terminale.

Les flux sortants Par rapport aux chiffres disponibles dans Parcoursup, 70 % des élèves ayant gardé la spécialité NSI en terminale avaient également gardé la spécialité mathématiques. Ils sont 12 % à avoir gardé une autre spécialité scientifique, et 6 % à avoir gardé la spécialité Sciences économiques et sociales (SES). La répartition des orientations choisies par les élèves ayant gardé la doublette Maths+NSI est la suivante :

BUT	Écoles d'ingénieurs	Licence Info.	BTS	Licence Math	CPGE
26%	16%	14%	9%	7%	6%

TABLE 2. Orientations des élèves avec Maths+NSI.

Dans le cas particulier de ces élèves ayant en plus pris l'option Maths expertes en terminale et ayant eu la mention Très bien (TB) au bac, la répartition est plutôt la suivante :

Écoles d'ingénieurs	CPGE	Licence Math ou Info	BUT
39 %	30 %	17 %	7 %

TABLE 3. Orientations des élèves avec Maths+NSI+ME et mention TB.

Le taux d'étudiants venant de NSI en classes préparatoires aux grandes écoles (CPGE) est relativement faible par rapport à ceux venant de physique-chimie, qui ont aussi plus de classes préparatoires accessibles. En mathématiques, physique, ingénierie et informatique (MP2I), la classe préparatoire créée pour les élèves ayant pris la spécialité NSI, les taux de recrutement M+NSI+ME / M+PC+ME sont dans un rapport 37 % / 63 % (il y a donc moins d'élèves provenant de NSI que d'élèves provenant de physique-chimie). Par contre, en cours d'année, quand les élèves doivent choisir entre renforcer l'informatique ou s'orienter vers les sciences de l'ingénieur, ils sont 90 % à choisir informatique.

CAPES NSI et Agrégation d'Informatique

Le CAPES NSI en est à sa troisième année, avec un nombre de places par an tournant autour de la cinquantaine (30, puis 60 et enfin 50). Les admis représentaient l'année dernière 26 % des présents à l'écrit.

L'agrégation en est à sa première année. Après les écrits, il y avait 55 admissibles pour 20 postes [4]. Le ministère n'a pas souhaité communiquer cette année sur le nombre de candidats aux différents concours de l'agrégation.

La structuration du paysage de l'enseignement de l'informatique à l'entrée dans le supérieur
État des lieux sur l'accueil des nouveaux bacheliers

Classes préparatoires MP2I et MPI

Roland Groz, enseignant-chercheur à l'ENSIMAG, qui a participé à la rédaction du programme d'informatique des nouvelles classes préparatoires MP2I et MPI, fait un état des lieux à la fin de cette première année de fonctionnement [5].

Tout d'abord, il rappelle que les ouvertures de MP2I se sont faites à effectifs constants, par transformation d'autres classes préparatoires (souvent des MPSI). Ceci a donc *de facto* freiné les proviseurs dans l'ouverture de MP2I. Par ailleurs, cela a implicitement amené à ce que ce soient surtout les gros lycées, avec plusieurs MPSI qui ouvrent ces nouvelles classes préparatoires.

Roland Groz rappelle également le but premier des MP2I : proposer un accès aux classes préparatoires à des élèves n'ayant pas gardé la doublette math-physique en terminale, même si le programme de physique de première reste *a priori* un prérequis. Cependant, beaucoup de lycées ne proposant pas la spécialité NSI, il a été prévu, dès le départ, que l'accès à la classe préparatoire MP2I serait également ouvert à des élèves n'ayant pas suivi la spécialité NSI. Cela semblait indispensable pour

4. https://agreg-info.org/.

5. Présentation accessible à https://www.societe-informatique-de-france.fr/wp-content/uploads/2022/09/SIF_mai2022_RG.pdf.

garantir le succès de cette nouvelle CPGE. Roland Groz précise également que cette ouverture représente une évolution importante dans un monde (celui des classes préparatoires) assez conservateur. Malgré tout, R. Groz rappelle qu'en terme de volume horaire, l'informatique reste la troisième matière après les maths et la physique.

Le fait que la spécialité NSI ne soit pas un prérequis a été une des raisons du choix des langages C et OCAML en MP2I plutôt que Python, pour limiter les écarts entre les élèves ayant suivi NSI et les autres. Une autre raison est liée à la structure générale des cours en classe préparatoire, où l'on préfère enseigner les fondements théoriques des disciplines, notamment car cela est plus facile à évaluer équitablement dans les concours.

Du fait de l'ouverture essentiellement dans de grands lycées et de la vision très hiérarchisée des écoles d'ingénieurs et des lycées avec classes préparatoires, le taux de sélection des MP2I a été l'un des plus forts ; cette nouvelle filière ayant par ailleurs été très attractive. Si une minorité d'élèves recrutés en MP2I avait gardé NSI en terminale, plus de 80 % avaient toutefois suivi la spécialité NSI en première.

Les MP2I/MPI représentent surtout un choix stratégique pour choisir les meilleures écoles d'ingénieurs, beaucoup d'écoles prestigieuses ayant ouvert un nombre important de places pour les élèves issus de ces CPGE, comme l'École polytechnique ou l'École normale supérieure. Plus généralement, au concours commun INP, les places offertes aux élèves de MP correspondent à 15 % des candidats, alors qu'il devrait y avoir des places pour 43 % des élèves issus de MPI.

En BUT

La situation pour les bachelors universitaires de technologie (BUT) informatique est présentée via deux expériences : celle de l'IUT de Lille, relatée par la cheffe du département informatique Marie Deletombe, et celle de l'IUT de Vannes, décrite [6] par Brigitte Le Pévédic, directrice de l'IUT de Vannes.

Les spécificités du BUT sont rappelées : il s'agit maintenant d'une licence professionnelle délivrée après trois ans d'études, avec un volume horaire en légère hausse par rapport à l'ancien DUT sur 2 ans (+10 %), avec de nombreuses heures de projet (600 h) et de stage (22 à 26 semaines). Les BUT proposent maintenant différents parcours (quatre pour le BUT informatique, deux pour le BUT statistique et informatique décisionnelle (proposé à Vannes), que chaque IUT décide de mettre en place ou non. En ce qui concerne le BUT informatique, les 46 départements informatique de France proposent au moins le parcours Réalisation d'applications : conception, développement, validation. L'évaluation est maintenant une évaluation par compétence, devant reposer pour environ 50 % sur les projets et les stages. Le recrutement en IUT doit maintenant offrir au moins 50 % de ses places aux élèves issus de bacs technologiques (STI2D essentiellement pour le BUT informatique). Dans la plupart des

6. Présentation accessible à `https://www.societe-informatique-de-france.fr/wp-content/uploads/2022/07/Brigitte_Le_Pevedic_BUT_2022.pdf`.

IUT, et c'est notamment le cas à Lille, cela conduit à une forte sélectivité sur les bacs généraux et par contre à un épuisement de la liste des candidats provenant des bacs technologiques, mais la situation est un peu différente à Vannes où, l'historique de l'IUT permet une forte attractivité des bacs technologiques, si bien que l'IUT réussit à recruter 50 % de ces élèves sans épuiser la liste des candidats. Concernant les élèves de bac général effectivement recrutés, une majorité avait pris la spécialité NSI au moins en première. Le faible nombre de filles demeure cependant un problème (8 %).

L'hétérogénéité des étudiants recrutés devient un problème difficile à gérer. En effet, les taux de réussite sont radicalement différents entre les élèves provenant de bacs généraux et ceux venant de bacs technologiques. Différentes solutions sont envisagées (tutorat par d'autres étudiants, semaines de rattrapage en début d'année, par exemple). L'opportunité d'établir des groupes de niveaux reste discutée selon les départements.

Concernant les poursuites d'études, il y a plusieurs possibilités : licence 3 ou école d'ingénieurs après le DUT (toujours délivré à la fin de la deuxième année), licence 3 ou école d'ingénieurs après le BUT, ou encore master 1 après le BUT. Le cas général sera probablement une poursuite à Bac+3 après le BUT, et des accords avec les universités et écoles d'ingénieurs locales pourraient mener, au cas par cas, aux autres possibilités.

En licence informatique

L'organisation des licences d'informatique n'étant pas définie par un programme national, contrairement aux BUT, deux organisations différentes sont présentées : celle de Bordeaux [7], par Laurent Réveillère, responsable de la licence informatique, et celle de Toulouse [8], par Olivier Gasquet, directeur du département informatique.

À Bordeaux, le choix fait a été de proposer des UE à deux niveaux, les étudiants intégrant un niveau ou l'autre suite à un test de positionnement. Cela permet d'éviter l'ennui aux élèves ayant suivi NSI, sans exiger pour le moment NSI comme un prérequis. Cependant, la forte attractivité de la licence et le taux important d'étudiants arrivant après avoir suivi au moins une année de NSI pourraient amener à revoir cette situation. À terme, cela pourrait même conduire à descendre d'une année certaines matières, et rehausser ainsi le niveau du master.

À Toulouse, un choix différent a été fait : chaque UE (sous la forme de cours–TD mélangés) est proposée aux deux semestres, avec des modules de remédiation pour les élèves n'ayant pas suivi NSI ou ayant des lacunes en mathématiques. Un élève peut suivre une UE à un semestre dès qu'il a validé les prérequis au semestre

7. Présentation accessible à `https://www.societe-informatique-de-france.fr/wp-content/uploads/2022/07/Laurent_Reveillere_Licence_Bordeaux_2022.pdf`.

8. Présentation accessible à `https://www.societe-informatique-de-france.fr/wp-content/uploads/2022/07/Olivier_Gasquet_Licence_Toulouse_2022.pdf`.

précédent. Il n'y a donc plus de redoublement général ni de compensation entre UE, mais cela permet aux étudiants d'avancer à des rythmes individualisés, avec une licence faisable en trois à quatre ans, voire en deux ans dans certains cas.

Expériences concrètes de méthodes d'enseignement de l'informatique dans le supérieur

La session de l'après-midi était dédiée à la présentation d'expériences pédagogiques mises en œuvre notamment dans le supérieur.

Informatique sans ordinateur

Dans un première temps, Pascal Lafourcade, enseignant-chercheur à l'IUT de Clermont-Ferrand, revient sur l'historique de l'informatique sans ordinateur (ISO) [9]. L'invention de l'ISO remonte à l'année 1992, où trois néo-zélandais (Tim Bell, Ian Wiiten, et Michael Fellows) ont créé le concept notamment pour pallier les défaillances réseau de l'époque. Cela a donné naissance au site *Computer Science Unplugged* [10]. En France, un groupe ISO informel s'est créé. P. Lafourcade liste plusieurs sites de ressources :

— le site Interstices [11] qui a traduit une vingtaine d'activités du site *CS unplugged* ;
— le site de Marie Duflot-Kremer [12] qui propose de nombreuses activités débranchées ;
— le site pixees [13] dont une partie est dédiée aux activités débranchées ;
— le site du groupe ISO de l'IREM de Clermont-Ferrand [14] ;
— les activités proposées par le groupe algorithmique de l'IREM de Grenoble [15].

Dans un deuxième temps, P. Lafourcade présente l'utilisation de plusieurs activités d'ISO, à différents niveaux. Certaines activités sont plus adaptées au primaire ou au collège. D'autres sont tout à fait adaptées à des élèves de NSI comme Mission cryptographie, par exemple, ou encore l'activité des Marmottes au sommeil léger de Marie Duflot-Kremer qui initie à l'algorithme de Huffmann pour la compression de données.

9. https://www.societe-informatique-de-france.fr/wp-content/uploads/2022/07/Pascal_Lafourcade_InfoSansOrdi_2022.pdf.
10. https://www.csunplugged.org/.
11. https://interstices.info/.
12. (https://members.loria.fr/MDuflot/files/mediation.html.
13. https://pixees.fr/.
14. http://www.irem.univ-bpclermont.fr/Informatique-sans-Ordinateur.
15. https://irem.univ-grenoble-alpes.fr/.

La pratique de l'ISO a également toute sa place dans le supérieur. Ainsi, P. Lafourcade l'utilise en BUT, activité sur les courbes de Béziers ; en licence professionnelle (LP) dans le cadre d'un cours de sécurité sur le web, et également en master 2, avec l'activité Antivirus, qui permet de comprendre l'indécidabilité du problème de l'arrêt.

Les activités types ISO sont en général très motivantes, et sont plus équitables pour les élèves ne jouant pas à des jeux vidéos (et notamment beaucoup de filles) qui, par conséquent, ne sont pas forcément aussi à l'aise avec le clavier que les autres. Par contre, pour que l'apprentissage cible demeure acquis, une phase de remédiation par l'enseignant à la fin de l'activité semble indispensable.

Caseine

Nadia Brauner présente ensuite Caséine [16]. Il s'agit d'une plateforme pour accompagner l'enseignement de l'informatique. Intégrée à Moodle, elle permet d'évaluer automatiquement du code informatique et des modèles mathématiques, d'assurer un suivi pédagogique interactif des étudiants, et de partager des ressources entre enseignants. Des *plugins* permettent d'interagir directement avec Caséine à partir de nombreux IDE.

L'évaluation individuelle formative motive énormément les étudiants, qui sont galvanisés par le fait d'améliorer leur score sans se sentir jugés par l'enseignant. Cependant, cela présente un risque dans la mesure où certains élèves finissent par passer trop de temps pour atteindre le score maximal, dont l'obtention peut quelques fois dépasser les objectifs du cours.

Méthodologie agile

Pour terminer la journée, Jannick Laval et Mathieu Vermeulen présentent le principe des méthodes agiles pour la pédagogie dans l'enseignement supérieur (ALPES) [17] qu'ils ont développées. L'idée de départ vient du constat suivant : en école d'ingénieurs, l'informatique ne motive pas beaucoup les élèves. Pour beaucoup, ils sont issus d'une formation maths-physique et veulent devenir manager. L'idée est alors, en reprenant les concepts du manifeste agile, de remettre l'humain (ici, l'étudiant) au centre, sans, bien sûr, oublier les objectifs pédagogiques. La méthode ALPES repose sur les fondements suivants :

— projet : pour motiver les étudiants, la réalisation d'un projet sert d'objectif ;
— *pair programming* : l'étudiant n'est pas seul face à un problème ;
— cours inversé : ainsi, les interventions de l'enseignant répondent à un problème précis auquel l'élève se trouve confronté ;

16. https://doi.org/10.48556/SIF.1024.20.107, https://moodle.caseine.org.
17. http://jannik-laval.eu/approchesalpes/.

— socio-constructivisme : l'étudiant doit être actif dans le processus d'apprentissage, mais la connaissance se construit principalement à la suite d'interactions sociales.

La mise en œuvre de la méthode ALPES, reposant sur la technique du Pomodoro [18] peut être résumée ainsi :

— un support de cours auto-suffisant est accessible en ligne ;
— une séance de cours est divisée en une succession de *longueurs* de 20 minutes de travail séparées par des pauses de 5 minutes.

Le respect des temps est fondamental. De même, les 20 minutes de travail doivent être 20 minutes de travail effectif, tandis que les pauses doivent être des moment de vraie pause. Lors du premier cours, les deux premières longueurs sont dédiées à la présentation du projet et à la mise en place des outils nécessaires à la suite. Un *planning board* préalable, qui liste les *User Stories* à implanter pendant chaque séance de cours (ou *sprint*) est établi. Les *User Stories* (concept repris de la méthode Scrum) sont des fonctionnalités à implanter. Elles doivent être indépendantes les unes des autres. Chaque *User Story* est découpé en tâches (le découpage peut être donné ou non). Des *Users Stories* plus ou moins optionnelles permettent de gérer les différences de rythme et de niveau des élèves.

Lors d'une séance, les élèves disposent d'un *Task board* : un tableau avec 3 ou 4 colonnes (*TO DO*, *DOING*, *DONE*, et éventuellement *HELP*) où des *post-it* associés à chaque tâche sont déplacés de colonne en colonne au fur et à mesure de l'avancée.

À la fin de la séance, les élèves remplissent un *Tweetback board* qui permet aux élèves de faire une introspection sur le travail réalisé et les notions acquises.

Conclusion

L'arrivée dans le supérieur des étudiants ayant suivi la spécialité Numérique et sciences informatiques est en train de changer le paysage de l'enseignement de l'informatique en France. De nombreuses adaptations, déjà initiées dans certaines formations, vont devoir être généralisées, non seulement pour gérer l'hétérogénéité du public issu du nouveau bac, mais plus spécialement pour prendre en compte les acquis de cette nouvelle spécialité. Face à ces changements, les enseignants du supérieur réagissent et des travaux sont engagés sur la pédagogie et la didactique de l'informatique.

18. `http://pomodorotechnique.com/`.

Accompagner l'enseignement de l'informatique avec Caseine

Astor Bizard, Denis Bouhineau, Nadia Brauner, Hadrien Cambazard, Nicolas Catusse, Mario Cortes Cornax, Aurélie Lagoutte, Pierre Lemaire, Yvan Maillot, Florence Thiard

C'est quoi cette bouteille de lait ?



Caseine avec son logo ci-contre est à la fois une plateforme d'apprentissage en informatique, mathématiques et génie industriel, et une communauté d'enseignants rassemblés autour de l'objectif commun de favoriser l'apprentissage en développant l'autonomie des étudiants et en améliorant la qualité du temps que l'enseignant leur consacre [1]. De sa création à aujourd'hui et au travers de ses évolutions, Caseine s'est toujours appuyée sur 3 piliers :

— des outils pour fournir un retour automatisé sur les productions des apprenants (code informatique, modèles mathématiques [1], ainsi que d'autres objets computationnels comme des circuits logiques, des jeux de test [2]...) ;

— un environnement permettant le suivi pédagogique des étudiants (Caseine est une instance de Moodle) ;

— le partage des pratiques et des contenus entre les enseignants, à travers des outils (espace de partage, visite de cours), des contenus et une communauté vivante d'utilisateurs (événements et outils de communication).

1. Un tour de la plateforme en 3 minutes : https://caseine.org/mod/page/view.php?id=19571.

DOI:10.48556/SIF.1024.20.107 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 107–122

La plateforme, hébergée aujourd'hui par le *datacenter* universitaire Grenoblois, est ouverte à tous et en particulier aux établissements d'enseignements supérieurs français. Elle a été utilisée par plus de 9 000 personnes pendant l'année universitaire 2021-2022, provenant essentiellement de 13 universités différentes (dont deux à l'international) et plus de 10 lycées répartis sur l'ensemble du territoire. Certaines institutions ont généralisé son usage (l'université Grenoble Alpes et Grenoble INP, à l'origine du projet, mais aussi l'université Clermont Auvergne), d'autres l'utilisent pour une partie de leurs enseignements (université de Haute Alsace, université de Bordeaux, certains lycées, etc.). Elle héberge également des projets universitaires, qui contribuent volontairement à son financement.

Est-ce différent de Moodle ?

Ni oui, ni non. Caseine est une instance de l'environnement numérique d'apprentissage Moodle, comme il en existe dans de nombreuses universités ; mais à la différence de beaucoup, elle n'est pas exclusivement adossée à une institution spécifique. Via Renater et plus largement la fédération EduGain, toutes les personnes disposant d'un compte universitaire français, en particulier, peuvent s'y connecter. Les enseignants d'établissements publics peuvent demander l'ouverture d'un ou plusieurs espaces de cours, aussi bien pour expérimenter, partager, que pour y accueillir leurs étudiants.

Certains cours sont en accès libre, en tant qu'invité ou sur auto-inscription. Une part de ceux-ci, mis en avant en tant que « cours ouverts » n'ont pas vocation à accueillir directement des promotions d'étudiants : ces espaces servent à l'autoformation (d'initiative individuelle ou en complément d'un cours), mais aussi à donner aux enseignants et enseignantes intéressés une idée des possibilités de la plateforme, les ressources qu'ils contiennent étant placées sous une licence autorisant leur réutilisation (généralement CC-BY-SA). On y trouve aussi bien des cours « clés en main » qui peuvent être dupliqués et utilisés comme base de nouveaux cours (initiation à la programmation en python, recherche opérationnelle, programmation objet...) que des espaces à vocation de démonstration (jeu de piste SQL, méthodes de test logiciel) ou de catalogue de ressources (vitrine pour l'enseignement de l'informatique en lycée). Les autres cours présents sur la plateforme, qui peuvent être en accès libre ou restreint, sont généralement gérés par des enseignants et des enseignantes et destinés à être utilisés dans le cadre de leurs enseignements.

Moodle offre aux enseignants un environnement complet pour assurer le suivi pédagogique des étudiants et suffisamment riche et souple pour scénariser finement leurs cours pour différents usages (cours complet ou ressources de compléments, préparation, entraînement ou approfondissement, en autonomie ou dans le cadre d'un

enseignement hybride...). Et surtout, cet environnement s'appuie sur une large communauté dynamique, et le cœur de la plateforme est largement extensible par des modules (types d'activités, outils de suivi...) développés par des institutions ou utilisateurs et publiés via le marché d'application Moodle. Côté Caseine, selon les besoins exprimés et discutés entre les enseignants, l'équipe sélectionne ceux qui s'avèrent pertinents. Si une fonctionnalité n'existe pas, le développement d'un module ad-hoc peut être envisagé, avec l'objectif de publier sur le Moodle Store les fonctionnalités généralisables et susceptibles de répondre aux besoins d'autres institutions. Les modules suivants ont ainsi été publiés par l'équipe Caseine [2] :

— le bloc « Point de vue », un outil permettant de recueillir sous forme de « réactions » le point de vue des étudiants sur la difficulté des activités proposées ;
— l'activité *Random Activity* — développée pendant le confinement — permettant d'attribuer aléatoirement des activités aux étudiants d'un cours (utile par exemple dans le cadre d'examens) ;
— les questions de programmation « Questions VPL » permettant d'intégrer de courtes questions de programmation dans des quiz Moodle ;
— le comportement de questions *student feedback* permettant de recueillir des commentaires étudiants sur les questions d'un test Moodle.

Des outils de retour automatique et d'aide à l'évaluation

Un des modules installés sur Caseine occupe une place particulière. Intégré dès les origines du projet, il représente une grosse partie des activités existantes sur la plateforme. Il s'agit du *Virtual Programming Lab* (VPL) (cf.figure 1), développé et maintenu principalement par Juan Carlos Rodriguez-del-Pino, de l'université de Las Palmas (Canaries). Ce module permet de réaliser des activités de programmation complètes sur Moodle, incluant environnement d'édition et d'exécution, ainsi que des outils de retour et d'aide à l'évaluation automatisée. La version présente sur Caseine diffère légèrement de la version canonique publiée sur le marché Moodle : nous y apportons des corrections et des « améliorations » (de notre point de vue !) ; toujours dans le même esprit, nous remontons et proposons l'intégration des fonctionnalités susceptibles d'intéresser un public plus large.

À quoi ressemble une activité VPL du point de vue de l'étudiant ? L'édition de code (éventuellement à partir d'un squelette fourni) s'effectue dans la partie centrale. Le bouton « fusée » permet d'exécuter le programme (en environnement textuel ou graphique selon l'activité). Le bouton « coche » permet d'évaluer le travail (en lançant les tests prévus par l'enseignant). Le rapport d'évaluation (résultats de tests,

2. `https://moodle.org/plugins/?q=caseine`.

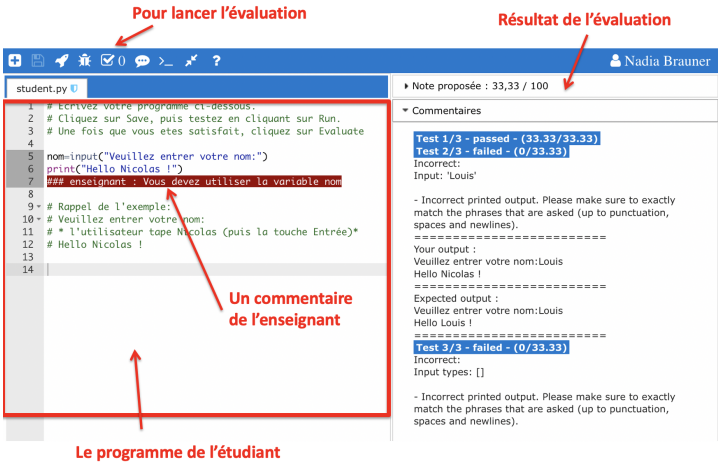


FIGURE 1. Une activité VPL du point de vue de l'étudiant.

commentaires sur les erreurs éventuelles...) apparaît alors dans le volet de droite, associé éventuellement à une note indicative.

L'enseignant, lui, a accès à toutes les soumissions de ses étudiants, ainsi qu'aux rapports d'évaluation et aux notes automatiques attribuées le cas échéant. Il peut évaluer les travaux, les modifier, les commenter, ajuster la note... Il a également accès à une série de statistiques, ainsi qu'à un outil de détection de similarité, permettant de repérer les productions d'étudiants potentiellement trop proches les unes des autres. Remarquons que chaque action « enregistrer », « exécuter » et « évaluer » de l'étudiant donne lieu à une soumission, visible par l'enseignant. Cela présente des avantages, permettant par exemple une aide « en direct » de l'enseignant (qui peut par exemple ajouter des commentaires spéciaux dans le code, que l'étudiant retrouvera dans son interface).

Comment les résultats d'évaluations automatiques sont-ils produits ? Le principe de fonctionnement du VPL est au fond très simple : lorsque l'étudiant lance une action « évaluer », l'ensemble des fichiers fournis par l'étudiant ainsi que certains fichiers fournis par l'enseignant sont envoyés sur le serveur d'exécution (un environnement GNU/Linux classique). La séquence d'évaluation (compilation, exécution des tests...) s'exécute alors dans cet environnement, produisant une sortie formatée, à partir de laquelle seront extraits le rapport d'évaluation à présenter à l'étudiant et la note obtenue. Même principe général pour une action « exécuter », sauf que la sortie produite est alors directement affichée dans la console du VPL.

Comment sont définies ces séquences d'exécution et d'évaluation ? Si l'enseignant n'a pas de besoin particulier et qu'il ne précise rien, le VPL propose un comportement par défaut. Pour l'exécution, il s'agira de compiler les sources et exécuter le premier fichier soumis par l'étudiant, et pour l'évaluation, d'un mécanisme type « boîte noire » : la sortie standard du programme étudiant est comparée avec une sortie attendue, pour une série d'entrées spécifiées par l'enseignant. Celui-ci doit alors simplement décrire une série de cas-tests (entrée, sortie attendue, message en cas d'échec...).

Si ce fonctionnement par défaut a pour avantage la simplicité et une relative indifférence vis-à-vis du langage de programmation, il peut en revanche ne pas suffire à l'usage envisagé (évaluations fonctionnelles, tests unitaires, évaluation de la complexité...). Il est alors possible d'écrire ses propres scripts (bash ou autre) et programmes d'exécution et d'évaluation : un script, un environnement complet ; la seule limite est l'imagination, ou presque ; on peut d'ailleurs tout à fait envisager des activités allant au-delà de la programmation au sens strict (calcul formel, statistiques...).

Ce dernier mode d'utilisation conviendra particulièrement, au moins dans un premier temps, aux enseignants rodés à l'évaluation semi-automatique, qui ont peut-être déjà leur propre « moulinette » (qu'ils souhaiteraient rendre plus accessible aux étudiants ou à leurs collègues), où à ceux qui ont une idée précise de ce qu'ils souhaitent faire et de la façon de le réaliser. Il demande toutefois un investissement conséquent en temps et en énergie. Le VPL prévoit heureusement un mécanisme d'héritage entre activités, permettant de partager les fichiers qui les composent (et en particulier les scripts d'évaluation). Ce mécanisme est mis à profit sur Caseine, en permettant d'utiliser directement et facilement des scripts d'évaluation développés par l'équipe ou la communauté et mis à disposition de tous (il suffit de choisir le mécanisme à utiliser lors de la création de l'activité, parmi les bases disponibles). Par exemple, on peut facilement créer des activités en Java ou Python en utilisant un framework de test unitaire (JUnit, python-unittest) pour spécifier l'évaluation, évaluer empiriquement la complexité des programmes étudiants, évaluer des requêtes SQL, ou encore des programmes étudiants modifiant l'environnement...

L'interface web de la figure 1 à l'avantage d'être simple, facilement accessible par les étudiants, sans installation nécessaire, et de les plonger directement au cœur de l'exercice. En contrepartie, elle déconnecte l'élève de l'environnement de développement et d'exécution sous-jacent à la programmation elle-même, et ne lui permet donc pas d'acquérir une pleine autonomie ni la maîtrise de ses outils. Elle est également peu commode pour la gestion de projets conséquents, et à l'autre extrémité du spectre, peut s'avérer trop bruyante pour des exercices « courts » (du moins dans leur expression). Il est bon tout d'abord de rappeler que le VPL est un outil d'aide à l'enseignement, avec ses limites, et a vocation à s'intégrer dans une scénarisation complète, mêlé à d'autres contenus, activités et modalités d'interactions. Toutefois, dans le cadre du projet Caseine, des interfaces alternatives ont été conçues qui peuvent

en élargir le périmètre d'usage. D'un côté, le module « questions VPL » permet de proposer une interface encore plus légère et épurée à une activité VPL, sous la forme d'une question de quiz Moodle. Cette forme est particulièrement adaptée aux exercices à réponses concises (quelques lignes de code, une requête SQL, une fonction mathématique...) ou qui s'intègrent dans un tout (un quiz Moodle). De l'autre, des plugins pour différents IDE (Eclipse, VSCode, IntelliJ...) permettent à l'étudiant de travailler (ainsi que compiler, exécuter...) en local dans un environnement de développement complet, et de soumettre et récupérer le rapport d'évaluation depuis leur IDE (voir section théorie des graphes pour une description plus détaillée et un exemple de contexte d'utilisation).

Une communauté d'enseignants et un espace de partage et d'expérimentation !

Un objectif majeur de Caseine dès sa création était de permettre le partage et la mutualisation, au sein d'une équipe pédagogique d'abord, puis plus largement d'une communauté d'enseignants, de ressources, de contenus mais aussi de bonnes pratiques et d'outils techniques. La plateforme se prête, de façon perméable, à toute une gamme d'usages s'inscrivant dans cet esprit de mutualisation, du terrain de jeu (mais pas seulement) individuel à l'expérimentation collective, à l'utilisation plus routinière de ressources éprouvées.

Par exemple, dans ses espaces de cours, un enseignant peut élaborer et essayer des outils d'évaluation qui conviennent à ses besoins, affiner ces outils et rôder des activités au fil des promotions et des retours étudiants, puis une fois ceux-ci arrivés à une maturité qu'il jugera satisfaisante, les partager avec la communauté : c'est en particulier ainsi qu'ont été élaborés certains mécanismes d'évaluation pour le VPL aujourd'hui disponibles pour tous ou les outils permettant l'utilisation de Caseine depuis différents IDE.

Dans tous les cas, l'appui de l'équipe Caseine peut également être sollicité pour, selon les cas, initier, accompagner ou stabiliser un développement, voire le prendre en charge entièrement.

Pour ce qui est des contenus eux-mêmes, deux modalités coexistent actuellement : le partage individuel d'activités (éléments du cours), ou la proposition d'un cours *ouvert*. Il est également possible de rendre son cours *visitable*, c'est à dire d'ouvrir un accès partiel aux enseignants de la communauté pour qu'ils puissent consulter (et récupérer) les activités partagées directement dans leur contexte (cf. figure 2). Précisons que dans tous les cas, la démarche repose sur une décision volontaire de l'enseignant auteur : rien ne l'oblige à ouvrir ses contenus au partage.

Ce partage de ressources constitue un attrait important pour les enseignants, d'autant que le coût d'entrée technique et l'investissement en temps nécessaire pour exploiter pleinement les possibilités de la plateforme et produire des activités est

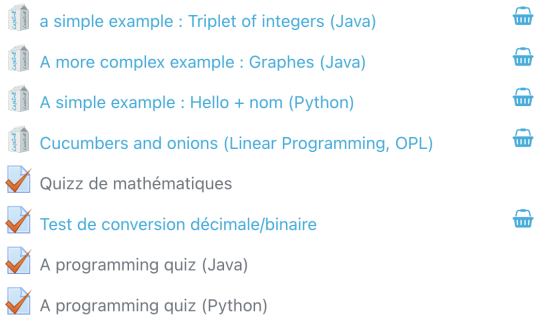


FIGURE 2. Exemple d'une section de cours visitable : les activités partagées peuvent être ajoutées au panier du visiteur qui peut ensuite les insérer dans son cours.

loin d'être négligeable. D'un côté, les nouveaux arrivants peuvent s'appuyer sur des contenus existants pour se mettre le pied à l'étrier, que ce soit en les intégrant dans leurs enseignements, en les utilisant comme base pour construire leur espace ou simplement comme exemple pour adapter leurs propres ressources. Pour les utilisateurs plus expérimentés, au-delà du gain de temps apporté par le réemploi de ressources, la perspective du partage ajoute un sens supplémentaire aux efforts de conception (on ne travaille plus uniquement pour soi et ses élèves, mais si on le souhaite, pour le bénéfice de la communauté) et conduit à terme à une amélioration de la qualité des ressources produites, via les relectures et retours des enseignants intéressés.

Toutefois, la mise en œuvre pratique est encore largement perfectible et en mutation ; elle est au centre de réflexions en cours pour, entre autres, affiner les outils, rendre leurs usages plus fluides, identifier la (ou plus sûrement les) granularité adéquate pour les ressources partagées, faciliter les retours et la collaboration.

Enfin, partager, c'est avant tout échanger, au-delà des outils et des modalités pratiques. Caseine est aujourd'hui aussi bien une plateforme qu'une communauté active d'enseignants qui partagent au sujet de leurs disciplines et leurs pratiques, au-delà du cadre de leurs équipe pédagogiques et institutions respectives.

La crise sanitaire de 2020 a marqué un tournant dans la fédération de cette communauté. En posant une série de défis techniques et pédagogiques (enseignement hybride, asynchrone, maintien de l'engagement des étudiants, évaluations à distance...) qui ont fortement sollicité la plateforme, les confinements et leurs suites ont été un déclencheur de discussions plus larges autour des usages numériques pour la pédagogie. Le chat (Mattermost) de la communauté Caseine, déployé à cette occasion pour leur offrir un espace, permet aujourd'hui d'échanger tant avec l'équipe qu'avec la communauté. D'un côté très pratique, il donne accès à un support technique collaboratif (de la part des ingénieurs Caseine mais aussi des utilisateurs experts), favorise

l'identification des besoins, le partage de compétences et l'émergence des nouvelles fonctionnalités ; mais il héberge également des discussions plus générales autour de la pédagogie, ainsi que des espaces de travail d'équipes pédagogique ou de projets hébergés sur la plateforme via des canaux thématiques publics ou privés (Ocaml, C, pédagogie, SQL...). Des webinaires « demi-heures Caseine », également mis en place au printemps 2020 sont aujourd'hui un rendez-vous semi-régulier, par visio-conférence, constitué d'un exposé suivi d'une discussion sur un sujet lié à Caseine... De près ou de loin ; il peut s'agir aussi bien d'une présentation d'une nouvelle fonctionnalité par l'équipe (« Le partage pour les nuls ou rendre son cours visitable »), d'un outil développé par un membre de la communauté (« Le plugin de conception C/C++ »), d'enseignant décrivant leurs usages (« Ludification sur caseine »), de sujets pédagogiques plus larges (« QCM augmentés », « évaluation par les pairs »)... Ces échanges sont enregistrées et rendus disponibles pour les enseignants, servant ainsi de base de connaissances.

Un commun numérique de fait

En fin de compte, depuis sa genèse et avant même la loi pour une république numérique en 2016, Caseine a inscrit la mutualisation — des ressources, des outils, des pratiques — au cœur de ses activités, ce qui améliore la qualité et l'efficacité tout en ajoutant du sens et du plaisir au travail fait. De ce fait, Caseine, en tant que projet, plateforme et communauté, s'inscrit naturellement dans la dynamique des communs numériques tels qu'ils sont maintenant définis et recommandés dans des textes officiels [3] [4]. Ainsi, le deuxième plan national de 2021 pour la science ouverte [5] stipule d'« *engager les acteurs de l'enseignement supérieur et de la recherche dans un travail commun sur les ressources éducatives libres pour les rendre plus visibles, mieux les partager et favoriser leur réutilisation* » ; c'est ce que s'efforce de faire Caseine depuis 7 ans.

Exemples d'utilisation de Caseine

Plus de 200 cours d'informatique ont été déposés sur la plateforme. La plupart sont de niveau universitaire, mais la plateforme accueille aussi des enseignements de niveau lycée (NSI, algorithmique). Issus des cours universitaires institutionnels, ils remplacent les sites institutionnels, tiennent lieu de site officiel pour des enseignements ou sont adossés à ces sites institutionnels. Le choix de cette plateforme externe peut étonner dans le cas où les plateformes institutionnelles existent, alors même que

3. https://labo.societenumerique.gouv.fr/2019/10/16/les-communs-numeriques-un-modele-innovant-de-developpement-des-ressources-numeriques.

4. https://communs.societenumerique.gouv.fr.

5. https://www.enseignementsup-recherche.gouv.fr/sites/default/files/2021-09/2e-plan-national-pour-la-science-ouverte-12968.pdf.

la multiplication des plateformes est régulièrement critiquée par les étudiants. Plusieurs explications de ce choix d'une plateforme externe à l'institution, stricto sensu, peuvent être données au vu des cours concernés.

Une part significative des cours disponibles sur Caseine sont en libre accès : soit ouverts aux anonymes, soit accessibles en auto-inscription, ce qui est plutôt l'exception sur les plateformes institutionnelles. Ce peut être une explication du choix de Caseine : la mise à disposition de contenus pour un plus grand nombre d'étudiants, le choix pour une plateforme plus orientée vers le partage entre enseignants, pour bénéficier des outils spécifiques mis en place pour favoriser l'ouverture et les échanges. Toutefois, l'ouverture et le partage ont des limites, les raisons du choix de Caseine ne peuvent s'y limiter.

Les cours disponibles sur Caseine sont rarement des cours « simples » (polycopiés et fiches d'exercices déposés sous forme de fichiers pdf, forum et une zone de rendu). Quasiment tous les cours d'informatique contiennent des activités de type VPL : avec en moyenne 16 VPL par cours pour l'ensemble des cours qui contiennent au moins un VPL, et 25 VPL pour les cours proposant 4 VPL ou plus. Le nombre de quiz est également significatif (bien que moins important que le nombre de VPL). Les cours proposés utilisent donc de manière significative des contenus orientés vers la mise en activité effective des étudiants sur la plateforme avec des retours automatiques possibles de leur production. D'autres modules sont également utilisés pour mettre en valeur les contenus et soutenir la motivation et l'activité des étudiants, mais les activités VPL dominent. La présence du module VPL, sa qualité intrinsèque, sa malléabilité ainsi que le support technique, pédagogique et à l'exploitation offert par la plateforme expliquent une grande partie des choix pour Caseine au lieu des plateformes institutionnelles. Ces dernières sont rarement aussi réactives et à jour des plugins de programmation (qui n'intéressent, par ailleurs, pas forcément beaucoup les autres disciplines, hors l'informatique) et des autres modules.

L'ouverture et le partage, le module VPL et le support pour les activités de programmation entraînent une présence massive d'auteurs actifs et ouverts, informaticiens pour la plupart, mais pas forcément tous intéressés par l'enseignement de la programmation, stricto-sensu. Cependant, la présence forte de VPL n'oriente pas nécessairement l'ensemble des activités vers la programmation. Des cours sur les couches basses de l'informatique (représentation de l'information, circuits logiques, langage machine, système), sur les données (SQL, SGBDR, conception de BD) et d'autres cours qui dépassent le noyau usuel de la programmation (Python, C, Java) sont également disponibles (langages exotiques, analyse syntaxique, test logiciel, complexité algorithmique, sécurité informatique, etc.). Dans la suite, voici quelques exemples d'usages variés du VPL et de leur intégration à Caseine.

Retour automatisé pour les requêtes SQL

La question revient régulièrement dans la communauté. Une première preuve de

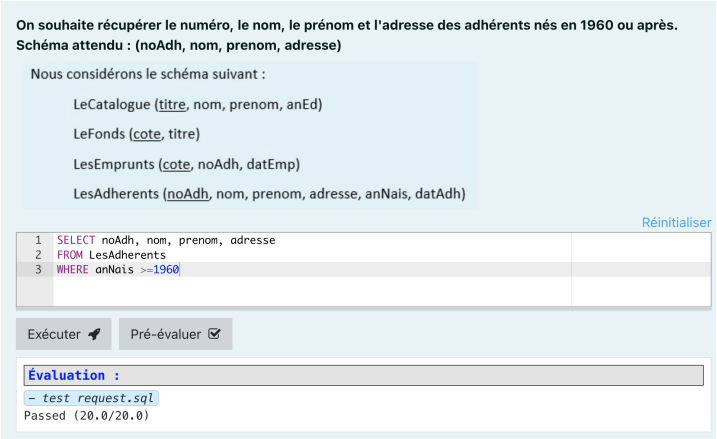


FIGURE 3. Évaluation de requêtes SQL.

concept et expérience à l'université Clermont Auvergne, un affinage du cahier des charges et un approfondissement de la réflexion par l'équipe pédagogique de l'université de Grenoble puis un développement soutenu par l'équipe Caseine aboutissent à un outil utilisé aujourd'hui par 2000 étudiants dans les deux universités, plusieurs lycées, un organisme de formation, ainsi qu'un cours « ouvert » en autoformation. L'idée initiale reprend l'esprit d'évaluation « boîte noire » du VPL : sans chercher à effectuer une analyse syntaxique approfondie de la requête soumise par l'étudiant, on se contente de l'envoyer au moteur de base de données (SQLite, serveur Oracle...) et de comparer les résultats obtenus avec ceux d'une requête de référence proposée par l'enseignant, sur une base de données idéalement suffisamment grande et variée (cf. figure 3). Ce choix à l'avantage de la simplicité et de l'indépendance vis-à-vis du dialecte, et l'inconvénient de la simplicité d'un retour « juste ou faux ». Pour détecter des erreurs courantes ou prédictives et fournir à l'étudiant un retour ciblé et des conseils de remédiation appropriés, on propose alors à l'enseignant de fournir non pas une mais plusieurs requêtes de référence, dont des requêtes incorrectes associées à un message personnalisé : la requête fournie par l'étudiant est alors évaluée en regard de chacune d'entres elles, jusqu'à trouver une correspondance, et c'est le retour (et la pénalité éventuelle) associé qui est appliqué. Notons qu'il n'est, en général, pas possible avec cette approche de distinguer deux requêtes donnant les mêmes résultats (même si des options permettent d'affiner, selon la présence ou l'absence de certains mots) : évaluer le style ou la performance reste la prérogative de l'enseignant. Cet outil a été utilisé à Grenoble pour transformer les travaux pratiques sur les requêtes SQL dans les enseignements de Licence en utilisant d'abord des VPL, puis des quiz Moodle avec questions VPL, interface qui se prête particulièrement

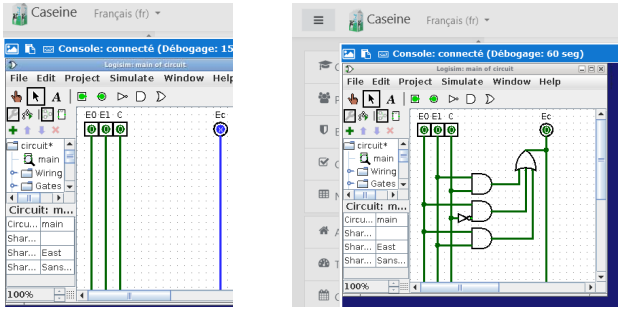
bien à cet usage (succession de petites questions à réponse courte), avec des effets positifs tant sur les suivis de TP que sur la correction de travaux notés, ainsi que sur l'intégration des ATER moins expérimentés dans l'équipe pédagogique. Pour les étudiants, comme pour les enseignants de passage, l'interface des questions VPL permet de masquer la complexité de l'environnement utilisé (Oracle), contourner les difficultés de connexion, et leur permet de démarrer rapidement en autonomie. Le suivi en temps réel de l'avancée des exercices pendant les TP s'est révélé utile (et pendant les périodes de confinement, indispensable) pour faire un support et des interventions personnalisées aux étudiants. Les retours automatisés (qui peuvent être adaptés en temps réel pendant la séance) favorisent l'autonomie des étudiants en TP, en débroussaillant les erreurs les plus classiques ; ce qui permet aux enseignants de se concentrer sur les difficultés de fond et le cœur de l'enseignement, ainsi qu'en facilitant la proposition de TP optionnels. Dans le cas des TP notés, le temps de correction a été réduit, les enseignants pouvant s'appuyer sur le retour automatique fourni dans l'interface de correction et y ajouter leur retour.

Circuits logiques

L'enseignement des circuits logiques (combinatoires, séquentiels, mémoires) pose de nombreux problèmes techniques (surtout) et pédagogiques. L'attrait de cette discipline vient de son aspect graphique, mais avant d'en faire un avantage, il faut pouvoir en surmonter les contraintes techniques : intégrer un outil graphique dans Moodle et VPL et adapter le processus de validation aux exercices de dessin de circuits.

Côté technique, le fonctionnement standard de Moodle/VPL repose sur les technologies web pour l'édition des programmes, l'intégration dans Moodle et sur un serveur Linux isolé (jail-server) pour l'évaluation des programmes et la mise au point (une interface Linux graphique est alors prévue). Deux pistes de solutions et deux technologies sont donc possibles pour l'intégration d'un outil graphique. En fait, les deux technologies vont être nécessaires. Parmi les logiciels de dessin de circuits logiques, c'est Logisim [6] qui a été choisi pour son ergonomie et ses fonctionnalités. Même si les sources de Logisim sont disponibles, l'intégration dans Caseine n'est pas passée par une adaptation du logiciel lui-même. En effet, l'exécution de Logisim sur le serveur Linux de VPL est facile. Fournir un circuit comme point de départ et permettre l'édition et la sauvegarde de ce circuit par l'utilisateur est aussi immédiat. Il reste à faire le lien entre Linux et Moodle. Le lien entre les fichiers de l'utilisateur sous Moodle et le serveur Linux isolé est prévu dans un seul sens par VPL (Moodle $\rightarrow$ Linux). Pour fournir un lien dans l'autre sens, un démon a été ajouté au serveur Linux pour scruter les changements de fichier utilisateur et envoyer via les services web Moodle/VPL prévus pour des soumissions à distance, les productions de l'étudiant. L'étudiant peut donc travailler avec Moodle/Linux/Logisim.

6. `http://www.cburch.com/logisim/`.



(a) (b)

FIGURE 4. Le circuit fourni initialement avec les entrées/sorties (a) et produit par l'utilisateur pour un multiplexeur (b).

Côté pédagogique, il faut aussi prévoir le test des circuits produits. À cet effet, Logisim a prévu une exécution des circuits à partir d'entrées fournies avec la production d'une trace des valeurs temporelles des sorties (afin de vérifier ces sorties). Le test prévu par VPL est similaire (tests basés sur un fichier d'entrées/sorties). Pour lier ces deux mécanismes de test, des transpositions sont effectuées des entrées VPL vers des fichiers d'entrée Logisim, puis après exécution de Logisim, des fichiers de sortie Logisim vers des sorties au format VPL. Pour aligner les entrées/sorties, la solution choisie fournit à l'étudiant un circuit initial ayant des entrées/sorties nommées suivant un ordre alphabétique spécifique. Remarque : pour les circuits séquentiels, c'est un peu plus compliqué, il faut ajouter ou prévoir un mécanisme (i.e. une partie de circuit supplémentaire) pour séquencer les entrées et mémoriser les sorties séquentielles.

L'intégration de Logisim à Caseine a mis en évidence plusieurs éléments. Le dessin de circuits logiques repose sur une application Logisim qui peut être transformée en activité (au sens de Moodle), sans changer l'application, à partir du moment où cette application peut s'apparenter à une boîte noire ayant des entrées et des sorties. Sur la base de ces entrées et sorties, une intégration à Caseine est possible, pour d'autres activités réductibles à une boîte noire similaire, en répétant le schéma suivi par l'intégration de Logisim par ajout de transpositions et d'un démon de communication. On peut ainsi songer à l'intégration d'activités de dessin d'automates formels, ou d'autres activités reposant sur des applications tierces graphiques ou non.

Pour les usages, le travail d'intégration et de mise en place d'activités de dessin de circuits a été accéléré par les conditions sanitaires de confinement. Cependant, la première année (2020-2021), le nombre d'activités VPL et l'intégration n'étaient

pas aussi avancée qu'aujourd'hui (juin 2022) ; mais les usages étaient plus importants (pour suppléer les présentiels manquants). En 2020-2021, les dessins de circuits faisaient partie des rendus du cours de L3 et ont été associés à la rédaction d'explications et à des calculs de complexité des circuits. Le contrôle de ces rédactions a été effectué par l'enseignant « à la main » et avec, hors plateforme, quelques outils en sus pour contrôler des aspects syntaxiques simples sur les circuits (vérification de l'absence d'utilisation inadaptée de la bibliothèque Logisim). À ce jour, le cours comporte 32 activités VPL, réparties en quatre groupes (premiers circuits, circuits combinatoires, automates et circuits à flots de données) ainsi que quelques vidéos. Plus d'une centaine d'étudiants de licence en ont profité, plusieurs centaines de circuits ont été dessinés. L'organisation en quatre groupes est liée de manière directe à la complexité des exercices. C'est par exemple, l'introduction des bascules ou points mémoires qui permet une complexité conceptuelle des circuits plus grande, mais qui est aussi associée à une complexité technique (le nombre de détails et de configurations possibles augmentent) et une taille de circuit aussi plus grande. En particulier, pour le passage des circuits combinatoires aux circuits séquentiels, le saut est important et l'effort pédagogique pour l'atténuer n'a peut-être pas été fait suffisamment dans ces activités VPL (pour le cours en présentiel, au tableau et avec papier-crayon, certaines séquences annexes rendent le passage aux circuits séquentiels plus progressif).

Une analyse informelle des productions des étudiants montre que l'activité est profitable à l'apprentissage en mesure de la complexité des circuits. Pour les circuits simples, le formalisme imposé de Logisim (plus rigide que celui demandé au niveau d'exercices papier-crayon) entraîne de meilleurs dessins : pas ou moins d'erreurs d'étourderie, qualité formelle des dessins, sans compter la clarté des tracés concrets (évidemment). Il semble qu'une partie des gains en apprentissage (en particulier ceux concernant la forme des dessins) a été conservée jusqu'à l'examen terminal. Cependant, pour les circuits plus complexes, la motivation étudiante et le nombre de solutions baisse avec Logisim, la proportion de situations d'évitement de la difficulté augmente (usage de circuits tout faits), même si l'activité reste motivante avec Logisim (comme l'ont fait remonter nombre d'étudiants). En comparaison avec le papier-crayon, les circuits séquentiels sont souvent incomplets et incorrects mais l'idée du circuit reste accessible.

Théorie des graphes

Les travaux pratiques de théorie des graphes peuvent prendre la forme de projets informatique conséquents qui nécessitent d'aller au-delà de l'interface web du VPL. Nous montrons l'usage du VPL dans le cadre d'un cours de théorie des graphes de licence d'informatique qui est dans la prolongation des cours de programmation et algorithmique. Ce cours partage de nombreuses activités avec les cours de recherche

opérationnelle. Pour un aperçu plus global, voir le cours ouvert et partagé de théorie des graphes [7] et [3] pour un aperçu plus détaillé.

Outre la maîtrise du domaine, ce cours vise, au travers de mises en activité sur machine, une prise de recul sur les structures de données utiles et sur la complexité algorithmique des solutions envisagées. Aussi, les programmes demandés, en Java ou en Python, sont basés sur une bibliothèque de classes fournie pour les implémentations élémentaires et si celle-ci fait quelques centaines de lignes, le travail étudiant ne concerne que les algorithmes de plus haut niveau qui ne font qu'une vingtaine de lignes.

Il s'agit, par exemple au cours d'une séance de TP de 3 heures, d'implémenter l'algorithme de Kruskal pour la recherche des arbres couvrants de poids minimum en utilisant une structure *Union Find* qu'ils doivent également programmer. Lorsqu'ils démarrent ce TP, les étudiants sont déjà familiarisés avec la classe `Graphe` fournie. L'algorithme de Kruskal a été vu en cours et en TD. Le TP doit permettre aux étudiants de consolider leur compréhension de cet algorithme et les forcer à réfléchir aux structures de données nécessaires pour une mise en œuvre effective et efficace. L'intérêt de la mise en pratique sur machine vient en particulier de cette confrontation à un objectif opérationnel qui ne peut être abordé avec autant d'acuité au tableau ou sur papier. L'utilisation d'une bibliothèque et la taille du code final complet permet aussi un travail à la frontière du génie logiciel et justifie l'utilisation d'IDE proposant des outils avancés de validation syntaxique, de documentation contextuelle, de complétion automatique, de débogage, etc. L'enjeu est alors de travailler sur de bonnes pratiques pour des codes qui commencent à devenir importants avec des objets complexes. En outre, certains étudiants utilisaient déjà ces IDE et l'équipe enseignante était favorable à une généralisation de ces usages. Cependant VPL, qui n'est pas un IDE, propose un environnement riche que l'on peut interroger de l'extérieur via des *webservices* ce qui permet les utilisations externes avec les IDE visés, VPL fournissant le lien avec Moodle et le système de validation qui lui est propre. Des *plugins* de connexion à Caseine ont donc été développés pour plusieurs IDE : VSCode/VSCodium, Eclipse, IntelliJ. Ces outils permettent de faire cohabiter l'environnement Moodle et un vrai environnement de développement : récupération (*pull* VPL→IDE) et envoi (*push* IDE→VPL) du code où le VPL permet de déposer les fichiers de l'étudiant et de faire tourner les tests automatiques. La connexion des étudiants à la plateforme est simple et transparente. Elle se fait au tout début de l'activité à partir d'un *token* d'identification Moodle de l'étudiant saisi une fois seulement par celui-ci avec le numéro de l'exercice VPL. Pour l'étudiant, c'est le seul lien avec Moodle, l'ensemble de son travail reste ensuite sur l'IDE ; le couple Moodle/VPL est complètement externalisé, il peut donc profiter et s'habituer aux IDE sans limite. Pour l'enseignant, le travail continue sur Caseine : il a accès aux

7. `https://moodle.caseine.org/course/view.php?id=147`.

codes et résultats des étudiants comme si ceux-ci étaient sur Caseine, l'utilisation de l'IDE est transparente pour l'enseignant.

Concernant la validation de l'activité, pour l'exemple Kruskal, l'activité fournit le code de base à l'étudiant et des tests fonctionnels, unitaires, de cas limites afin de permettre à l'étudiant de vérifier de façon incrémentale son avancée dans le travail demandé. En TP, généralement à la demande, l'enseignant peut lire le code, le valider ou le commenter et accompagner l'étudiant depuis Caseine en s'appuyant aussi sur les tests. Même si l'étudiant est incité à écrire quelques tests qui vérifient que l'algorithme fonctionne sur les cas « normaux », il ne lui est pas demandé de tester les cas particuliers même s'il doit les implémenter. Les étudiants voient une partie des tests qui seront utilisés pour évaluer leur travail. Ils ont ainsi accès à des tests correctement écrits dans un langage dédié au test (`JUnit` pour Java ou `Doctest` pour Python).

Un outil de retour automatique sur la complexité temporelle d'un algorithme. Comment réaliser automatiquement un retour à un étudiant sur la complexité temporelle de son algorithme ? Le problème est indécidable (au moins aussi difficile que le problème de la terminaison d'un programme) mais les complexités algorithmiques typiques rencontrées dans un cours d'informatique sont souvent assez limitées et se rencontrent au départ dans des cas académiques très épurés (recherche dans un tableau, recherche dichotomique, tris quadratiques, etc.). L'enjeu pédagogique de ce retour automatique est d'inciter un étudiant à aller au-delà « d'un programme qui fonctionne, qui passe les tests de validité ». Il doit permettre de provoquer des questionnements chez l'étudiant, sur la complexité algorithmique en général et sur celle de son algorithme dans un exercice précis. L'étudiant peut ainsi être sensibilisé à cet aspect et revenir vers l'enseignant.

L'outil développé comme une extension du VPL estime la complexité à partir du temps d'exécution de l'algorithme de l'étudiant sur différentes tailles de données d'entrée. Un ensemble de complexités possibles classiques (1, n, n^2, k^n, $n\log(n)$, 2^n...) sont comparées par régression linéaire ainsi que la complexité attendue et spécifiée par l'enseignant pour l'exercice. Les coefficients de corrélation et leurs intervalles de confiance sont utilisés pour éventuellement identifier une « bonne hypothèse ». En cas de succès de cette étape, le système fait un retour positif ou négatif (complexité meilleure ou moins bonne que celle attendue) à l'étudiant en indiquant la complexité identifiée. En cas d'échec, une régression non linéaire (de type $a+bx^c$ ou $a+be^{cx}$) est réalisée. Le système fait alors un retour positif ou négatif prudent sur la complexité observée en pratique et sa comparaison à celle attendue. Enfin, si la régression non linéaire colle mal aux données de temps CPU, l'analyse de la complexité est déclarée « non-concluante ». Cinq réponses sont ainsi envisagées : positive, négative, positive prudente, négative prudente, analyse non concluante. L'analyse doit être faite en ligne, rapidement et la taille des données à tester varie considérablement d'un exercice à l'autre. La taille maximum pour laquelle la limite de temps

du code étudiant est atteinte est rapidement identifiée afin de générer des tailles intermédiaires bien réparties dans cet intervalle dans le temps restant. Cette approche expérimentale ne permet pas de distinguer un facteur logarithmique dans les complexités (des données de taille exponentielle seraient nécessaires) et cette limite est indiquée dans le retour à l'étudiant. Pour utiliser ce système, l'enseignant doit spécifier une complexité théorique attendue pour un exercice mais aussi une méthode pour générer des jeux de données de taille paramétrée. Ces jeux de données définissent le type de complexité analysée (au pire cas, meilleur cas ou encore en moyenne) et exigent une réflexion soigneuse de la part de l'enseignant. Même si la mise en place pour l'enseignant nécessite quelques ajustements au début, par exemple pour le choix de la bonne famille de cas de tests générée, en pratique, cet outil fonctionne bien et permet vraiment d'accompagner les étudiants vers les objectifs de prise de recul sur les choix d'implémentation et les répercussions sur l'efficacité. Il a été utilisé entre autres dans des cours de recherche opérationnelle ou de théorie des graphes (voir ci-dessus) sur des centaines de codes d'étudiants. En pratique, il a par exemple permis dans ces cours d'identifier des codes exponentiels contre pseudo-polynomiaux en programmation dynamique, d'alerter les étudiants sur des récursivités avec mauvaise gestion de la mémoire, d'attirer l'attention des étudiants sur la complexité de leur structure *Union Find* pour l'algorithme de Kruskal, etc.

Pour rejoindre cette aventure, rendez-vous pour une première visite sur le site de Caseine [8].

Références

[1] Hadrien Cambazard, Nicolas Catusse, Nadia Brauner, and Pierre Lemaire. Teaching OR : automatic evaluation for linear programming modelling. 20(2) :333–345, 2022.

[2] Lydie du Bousquet and Christophe Saint-Marcel. De l'adaptation de caseine pour l'évaluation des tests des étudiants. In *21èmes Journées Approches Formelles dans l'Assistance au Développement de Logiciels (AFADL)*, 2022.

[3] N. Catusse, H. Cambazard, N. Brauner, B. Penz and F. Fontan. Innovative ideas for teaching supports : Application to Graph theory. arXiv :2209.05078, 2022.

8. `https://moodle.caseine.org`.

FICTION

La ronde des avatars

Serge Abiteboul [1]

« Votre méta est fait pour vivre et pas seulement pour être nettoyé. Offrez-lui CIF ! ». Le blond qui explique cela passe sa serpillère avec beaucoup de conscience sur le cyberespace.

J'allonge légèrement les jambes, je me cale dans mon fauteuil et je savoure une gorgée de mon verre de ratafia. Ça devient n'importe quoi les congrès de la SIF. Une pub de détergent maintenant pour renflouer les caisses. Un congrès de la SIF, un de plus. L'occasion de faire le point sur le domaine et surtout de retrouver des copains.

Je suis avec Maurice [2] qui tient la grande forme. En quelques minutes, il a démoli la dernière réforme de l'Éducation nationale et massacré les candidats aux élections présidentielles. Caustique, rafraichissant.

La terrasse du Dellys a été reconstituée avec un souci de détails depuis la table en formica jusqu'à l'assiette de « cahouètes ». Dans ce bar improbable, entre la gare de l'Est et du Nord, je me souviens avoir siroté *in real life* un ratafia avec lui. Nous avions parlé de patrimoine industriel, un intérêt partagé avec lui en dehors de l'informatique.

Il allume une cigarette. Des « *No smoking* » en lumière rouge se mettent à clignoter de tous les côtés. Mais, ils sont cons ! Qui est gêné par sa cigarette dans un monde virtuel ? Encore une règle états-unienne. Encore heureux qu'ils tolèrent le jaja. Maurice n'a rien remarqué. C'est dans ces moments-là que je regrette d'avoir arrêté de fumer.

Le haut-parleur sonne la fin de la session. Les congressistes se dispersent dans les grands salons du métavers. Quelques anciens nous font un signe amical de loin.

1. Directeur de recherche Inria et membre de l'Académie des sciences.
2. Maurice Nivat, https://fr.wikipedia.org/wiki/Maurice_Nivat.

DOI:10.48556/SIF.1024.20.123 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 123–124

Maurice est resté populaire. Une bande de jeunes très joyeux passe près de nous ; ils préparent une méta-soirée où nous ne serons pas invités.

Maurice éclate de rire :

« *Leurs boums virtuelles, question sexualité, ça ne vaut pas tripette.* »

Un ange passe. Le souvenir des boums d'antan ?

Quelque chose me tracasse. Est-ce de parler avec Maurice alors qu'il est mort depuis longtemps ? Son avatar est assez convainquant. Non. On s'est habitués à ce que des disparus participent au métavers même si certains disent que cela finira par nous glisser des hannetons dans le ciboulot. On n'est pas idiots ; on sait bien qu'on est dans un monde virtuel, que si on croise Harry Potter ou Wonder Woman, c'est pour du beurre, que Maurice est décédé.

Alors quelle est la cause de mon embarras ?

Est-ce parce que je suis également mort depuis des années ? Non. Ils ont dû récolter assez de données sur moi pour créer mon avatar. Je me trouve d'ailleurs plutôt réussi, plus sympa qu'en vrai peut-être, même s'ils auraient pu me donner quelques années de moins.

Mais alors qu'est-ce qui me gêne ?

Ce qui cloche, c'est le Dellys. Comment ont-« ils » su que Maurice et moi avions pris un verre au Dellys ? Mon avatar vient de vérifier sur la toile. Pas la moindre trace de cette rencontre. Personne ne savait. Pourtant, « ils » ont su. Saint Nick Bostrom priez pour nous : notre monde n'est qu'une simulation informatique ? Le complot est suprême.

Au secours ! Il doit y avoir une explication plus simple, plus rationnelle.

Le nombre de verres de ratafia sur la table peut-être ?



FICTION

Mon rosbif est biaisé

Serge Abiteboul [1] et Julia Stoyanovich [2]

Une discrimination systématique par un algorithme est appelée un *biais*. Les biais peuvent avoir différentes origines, et ils correspondent tous à des situations de la vie réelle. Pour illustrer cela, pensez à cuisiner du rosbif.

Pat décide de publier un site internet de recettes familiales en commençant par la plus populaire : le rosbif. La mère de Pat, sa sœur et son frère préparent tous le rosbif à peu près de la même manière : « comme le fait grand-mère ». Ils utilisent tous la même recette — le même algorithme. Leur choix de cuisiner le rosbif de cette manière montre une sorte de biais de *familiarité* — privilégiant les goûts et les expériences familiers par rapport aux nouveaux.

Lorsqu'elle sélectionne une recette de rosbif familière de manière biaisée, Pat ne fait rien de mal — elle suit simplement une tradition familiale agréable. Mais dans d'autres cas, le biais d'habitude peut être nocif : il peut conduire à une discrimination illégale dans l'embauche, le prêt ou l'octroi d'une libération conditionnelle. Aussi inoffensive soit-elle, la recette de rosbif de la famille de Pat nous permettra de discuter différents types de préjugés.

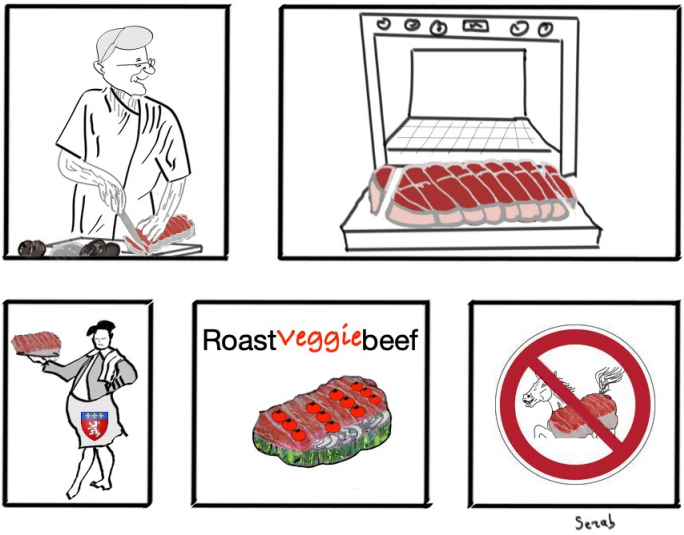
Comment choisir une recette de rosbif à suivre ? Commencez par sélectionner vos données d'entraînement, une recette « générique » que vous avez vu d'autres utiliser, et faites-la vôtre ! Dans le cas de Pat, le type de préjugé le plus visible vient des personnes dont elle adopte la recette : sa famille. Par exemple, au Maroc, la recette générique serait de préparer du bœuf avec du curcuma, du gingembre, du cumin et de la coriandre, ainsi que du jus de grenade. Par contre, en Alsace, la base du *ross*bif

1. Directeur de recherche Inria et membre de l'Académie des sciences.
2. Professeure à New York University.

DOI:10.48556/SIF.1024.20.125 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 125–127

n'est pas du tout du bœuf, mais de la viande de cheval (*ross* signifie cheval en alsacien). En clair, selon les personnes qu'on choisit, on obtient des recettes différentes. La famille de Pat est plutôt traditionnelle, ils utilisent du bœuf et le garnissent de persil, de sauge, de romarin et de thym (comme dans la chanson [3]). Le biais introduit par la sélection des données d'entraînement est parfois appelé biais *préexistant*, car il provient de recettes qui existaient avant que vous n'envisagiez même de faire vous-même ce rosbif.

Une pratique s'avère assez particulière dans la famille de Pat : ils découpent des tranches aux deux extrémités du rôti avant de le mettre au four. Pourquoi font-ils cela, demande Pat ? Un cousin propose : à cause de l'oxydation du rôti aux extrémités. Un autre n'est pas d'accord : parce que nous, les enfants, aimons notre rosbif bien cuit, et les petits morceaux sont alors hyper cuits. Pat demande à sa mère, sa tante et son oncle qui répondent : nous faisons ceci parce que maman faisait cela. Alors, quand elle en a l'occasion, elle demande à sa grand-mère qui répond : ma mère faisait comme ça. Mais pourquoi ? Parce que son four était trop petit pour contenir tout le rôti. C'est ce qu'on appelle le biais *technique*. En informatique, un tel biais peut être introduit, par exemple, parce qu'une représentation incorrecte des données a été choisie. Et ce n'est peut-être pas si facile à détecter car les programmeurs ont toujours tendance à supposer la perfection de leur code. Ou, du moins, ils remettent rarement en question un morceau de code qui fonctionne.



3. https://fr.wikipedia.org/wiki/Parsley,_Sage,_Rosemary_and_Thyme.

Pour introduire le dernier type de biais, revenons à l'Angleterre médiévale[4]. Il était une fois une ville avec deux chefs qui avaient inventé des recettes de rosbif concurrentes. Le duc qui avait entendu parler de ces deux nouvelles recettes se déplaça pour venir les déguster. Malheureusement, l'un des deux chefs eut un rhume ce jour-là. Ainsi, seul Maistre Chiquart, un chef français originaire de Lyon, put faire déguster son rosbif. Le duc l'apprécia tant qu'il l'introduisit à la cour du roi. Ce fut le début de la saga de la recette du rosbif que l'on connaît aujourd'hui, alors que la recette de l'autre chef se perdit dans l'histoire, jusqu'au nom de ce chef. Sans ce rhume ce jour fatidique, la recette de rosbif de Pat aurait peut-être été très différente. C'est ce qu'on appelle le biais *émergent* : plus les gens utilisent la recette de Chiquart, plus elle deviendra populaire et plus les gens l'utiliseront. Cela rendra la préparation du rosbif de plus en plus uniforme.

Le site web de Pat n'est pas très populaire auprès des *millenials*, la génération Y ; beaucoup d'entre eux sont végétariens et se soucient peu de recettes de rosbif. Un biais *générationnel* ?

4. Attention : cette histoire est une pure fiction. Il n'y a aucune preuve réelle que le rosbif ait été inventé par un chef français, même si cela est fort probable.

FICTION

Cours toujours

Fabien Gandon [1]

« *Pour comprendre la récursivité,*
il faut avant tout...
comprendre la récursivité. »
— Stephen Hawking

Mes anciens collègues sont partis et je reste assis dans ce rocking chair qu'ils m'ont offert ; prostré entre photo d'adieu et cliché de retraite. Mais il était temps que l'on me retire la craie et l'éponge car je commençais à les confondre avec le sabre et le bouclier. Ne vous méprenez pas, je ne compte pas encore d'Alzheimer dans mon carnet de contacts. J'étais juste devenu trop critique pour une formation uniforme, trop singulier pour l'éducation en série, et trop amer pour accepter sans mot dire la suite des évènements.

Hmmm...

Série et suite... Deux petites madeleines pour le « prof. d'info. » que j'étais. Et dire que Proust, dans ses brouillons, parlait d'une vulgaire tranche de pain grillé... ce carré rassis et fruste incomparable aux belles courbes lisses des madeleines.

> « *Voyez-vous* », m'avait dit l'inspecteur à sa dernière visite « *vous êtes à cet âge où vous avez l'expérience qui légitimerait que l'on vous donne les moyens de changer le système ; mais vous êtes aussi trop vieux pour être... comment dire...* contrôlable. »

1. Directeur de recherche, Inria Sophia Antipolis, responsable de l'équipe Wimmics (Université Côte d'Azur, Inria, CNRS, I3S), représentant Inria au W3C (*World Wide Web Consortium*), Fabien.Gandon@inria.fr — http://fabien.info.

DOI:10.48556/SIF.1024.20.129 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 129–138

Je n'avais rien trouvé à répondre au cynisme de la langue administrative qui me parlait, peut-être aussi par lâcheté puisqu'elle allait de pair avec la main qui me notait une dernière fois et fixerait, dans un jeu de dominos de formalités, le montant de ma retraite. Une dernière note, la note de bas de page, la page que l'on tourne.

Si vous m'interrogiez sur les raisons de mon amertume, j'ouvrirais un inexhaustible filon de critiques. Mais, pour faire court, j'ai acquis la certitude que l'antonymie des termes « programmes » et « éducatifs » comme nous les pratiquons maintenant, découle essentiellement du fait que nous gavons les masses estudiantines de prêt-à-penser en espérant qu'elles développent par magie leur propre mode de pensée. Même en cours de philosophie, nous leur assénons des vérités formolées, dans des débats *pro forma* à lire pour lundi prochain, pour leur reprocher ensuite de ne pas avoir de sens critique. Un formatage systématique, comme si la connaissance était un plat qui se mange froid et sur le pouce.

Dans notre enseignement à la chaîne, il n'y a plus de place et surtout plus de temps pour l'expérience intime du jugement personnel. Cette petite extase à l'instant d'un nouveau discernement. La compréhension, avec son petit *shoot* de dopamine pure injectée au cœur du cerveau. Nous n'apprenons pas à savourer l'entendement ; après le *fast-food* nous avons créé le *fast-knowledge*, et donc après la *malbouffe*... la *malpensée*. Je ne pense pas qu'il faille tout accélérer ; les temps de réflexion existent pour de bonnes raisons. À cela s'ajoute le fait que, dans nos programmes éducatifs, la mémoire remplace souvent la compréhension ; l'appris a remplacé l'acquis ; nous programmons l'élève au lieu de son épanouissement. Comme un tableau d'Arcimboldo : la forme y est, mais au lieu d'une toile de connaissances délicatement tissée, la substance est le patchwork abscons du prêt-à-savoir. Franchement, c'est un miracle que l'imagination survive à cette broyeuse de créativité.

Moi ? Amer ? Oui, vous avez raison. Mais avouez que c'est à n'y rien comprendre : lorsque l'on considère que nous sommes incontestablement entrés dans une ère où les réseaux d'information sont omniprésents, s'il est bien une fonction humaine dont ces réseaux sont la prothèse c'est la mémoire et s'il est bien une fonction que nous devrions développer en regard, c'est la critique et l'analyse. Il est moins urgent d'apprendre à mémoriser que de cultiver la lecture et la pensée critiques.

Pour parachever ce « tableau noir », certains s'étonnent aussi que notre système produise des individualistes. Mais que sont une note et son bulletin, une moyenne générale et son classement, une diplomation et son major sinon des incitations formelles à l'individualisme ? Et ceci dès l'entrée à l'école. On devrait médailler les humanistes et les esprits d'équipe qui survivent à vingt années de système éducatif. Comme Hugo le fit dire à Pan, « *tout le mal vient de la forme (...) on fait du ténébreux avec du radieux.* »

Vous voyez, il était temps que l'on me retire la craie et l'éponge et que quelqu'un prenne ma suite.

Hmmm...

Série et suite... Mes « *math'leines* » de Proust en effet, une pépite dans mes souvenirs de toutes ces années d'enseignement. J'inaugure une nouvelle espèce de dinosaures : je suis l'un des premiers professeurs agrégés d'informatique à prendre sa retraite. Mais si vous m'accordez un instant de votre vie, je repousse de quelques minutes cette retraite et je vous donne un dernier cours. Je vous ramène au début de ma carrière et je vous parle de cette élève qui est maintenant l'une de mes nombreuses collègues quelque part en France et à laquelle je pense depuis que je me suis assis dans ce rocking chair.

* * *

Suite et série... Un commencement est un moment d'extrême délicatesse comme l'écrivit Frank Herbert. Il y a quelque chose de terrifiant et de vertigineux à considérer tout ce qui peut dépendre des conditions initiales... comme, par exemple, les répercussions des premiers jours sur toute une vie ou, plus à propos, la rencontre avec un professeur qui sera en charge de votre premier contact avec une matière. Sachez donc, qu'au début de ma carrière, j'étais un pur produit de cette institution nationale de bourrage des crânes ; tellement bien démoulé que j'étais passé de l'autre côté de l'autel pour prêcher à mon tour la bonne pensée. Et paradoxalement, après toutes ces années d'études, j'allais enfin me prouver par l'exemple que l'un des meilleurs moyens d'apprendre est d'avoir à enseigner. Je ne dis pas cela uniquement parce que ce que l'on conçoit bien s'énonce clairement, mais aussi parce que les meilleures leçons restent celles de la vie.

Imaginez. Il était 12h15. La cloche issue, nul n'en doutait, d'un lot d'alarmes de casernes de pompiers, signifia si délicatement la fin des cours qu'elle résumait à elle seule le respect de toute l'institution pour son corps enseignant et ses élèves. Je posai la craie sur le bord du tableau alors que la salle se vidait dans le brouhaha des froissements de sacs en toile synthétique, des stylos s'entrechoquant, des fermetures éclair brusquées, des chaises raclant le linoléum avec empressement, des piétinements, des conversations privées et ponctuées d'éclats de voix et de rires, d'interjections et d'interpellations. C'est un rituel où pendant quelques secondes je prenais avec sévérité toute l'ampleur de la simulation de l'attention des élèves, juste avant de me rappeler que, sur les bancs quelques années plus tôt, je ne me comportais pas mieux.

Cela faisait maintenant deux ans que j'enseignais les mathématiques et l'informatique au lycée Pothier d'Orléans. Plus précisément, j'étais fier d'être parmi les premiers à avoir bénéficié de l'arrêté du 17 mai 2021 qui inscrivit au Journal Officiel la création de l'agrégation d'informatique après celle, en 2020, de la section « numérique et sciences informatiques » du Capes. Je constatai avec résignation, qu'une fois de plus il y avait plus de craie sur mon pantalon et sur mes mains que sur le tableau mais je continuais à préférer cette salle ; l'une des dernières à n'avoir pas abandonné le tableau noir et sa craie pour un tableau blanc ou un tableau numérique. À chacun

ses TOC : certains s'obstinent à porter la cravate, d'autres restent des inconditionnels de la craie. Alors que je me retournai vers mon sac, je vis qu'une de mes élèves attendait à coté du bureau. Un sourire torve sous un regard condescendant, comme on les réussit si bien à l'adolescence, exprimait un verdict sans appel sur ma maîtrise du calcaire blanc.

Stéphanie Ga.... Gral... Glam.... Glarenière ! Stéphanie Glarenière. Je la regardais, l'invitant par mon insistance et un haussement de sourcils à expliquer ce qu'elle attendait.

> — « *Je n'ai pas très bien compris ce que vous vouliez dire dans le 'grand 1' du chapitre 5 'récursivité et récurrence'* » commença-t-elle sans préambule et avec au regard l'agressive assurance qui la caractérisait.
> — ...

Je ne dis rien. Le silence est souvent une très bonne réplique initiale dans la maïeutique d'un dialogue pédagogique. C'est une invitation libre au développement et à l'approfondissement... mais je m'égare et, pour être honnête, d'un point de vue plus terre à terre le silence permet surtout de voir où l'autre veut en venir. Cette brillante élève était coutumière des séances vindicatives d'après cours où elle venait régler ses comptes avec ma pédagogie balbutiante. Et je savais déjà que le silence m'en dirait plus sur là où elle voulait vraiment en venir que toute question de ma part. C'est amusant de voir que le silence dans une conversation est si mal vécu dans nos sociétés occidentales occidentales là où, en orient, il est souvent la marque d'un temps de réflexion très respecté. Ici, il faut au plus vite le briser alors que là-bas il est précieux, cultivé et respecté. Mais laissons passer les anges et écoutons ce petit démon.

> — « *Là !... voilà !* » dit-elle en se raidissant le doigt planté sur ses notes « *Vous dites qu'une suite peut être entièrement définie par la valeur de son premier terme ou condition initiale... sa première valeur quoi...* » elle ponctua son interjection d'un haussement de sourcil circonflexe qui accentuait tout le bien qu'elle pensait de ma formulation « *et par une relation de récurrence, c'est-à-dire une relation qui définit un nouveau terme de la suite, en fonction d'un ou plusieurs termes précédents* ». Elle haussa les deux sourcils à l'horizontale dans une profonde expiration pour souligner un soupir très explicite « *Par exemple... je pars du terme 1 et je dis que je multiplie le terme précédent par deux à chaque fois et j'obtiens 1, 2, 4, 8, 16, etc.* »
> — ...
> — « Puis vous embrayez en disant, je cite, qu'un algorithme récursif résout un problème en s'appelant lui-même pour calculer des solutions plus petites... »

Sur cette phrase, elle laissa tomber sa mâchoire en biais sur la gauche, baissant la tête et, dans un même geste, levant les yeux sur moi et au ciel pour exprimer une exaspération qu'un metteur en scène aurait qualifiée de surjouée, a minima. Mais

l'adolescence n'est-elle pas un grand surjeu de la vie pour mieux se préparer à la seule et unique représentation qui nous est offerte ?

Devant ce silence qui lui paru de trop, elle explosa... enfin.

> — « *Franchement* » accusa-t-elle en relevant la commissure droite de ses lèvres « *Comment on peut imaginer quelque chose qui se répète de façon infinie en s'appuyant sur lui-même ?! Ça n'existe pas ! C'est encore un truc complètement théorique !* »

Nous y étions. Là, j'aurais dû reprendre la formulation un peu obscure et l'éclaircir à la craie blanche. J'aurais dû m'asseoir, prendre ses notes, reformuler, traiter un ou deux exemples, puis revenir à l'idée générale par abstraction.

Mais quelque chose venait de se libérer dans mon cerveau. Comme une veine d'idées qui se serait ouverte et se déverserait dans une hémorragie d'intellection. Je savourais l'exquise sensation que l'on ressent lorsque l'on atteint une nouvelle compréhension, ce profond frisson de contentement qui nous parcourt lorsque nous faisons l'expérience kinesthésique d'une extension de notre entendement. Je me sentais loin des techniques pédagogiques qui demanderaient d'illustrer par l'exemple et cependant, à bien y réfléchir, j'en étais peut-être plus près que jamais.

L'étudiante arborait un sourire emprunt de moquerie, d'inquiétude et de doute devant mon air ahuri. Je lus dans son regard d'adolescente une oscillation entre la peur d'être témoin d'un AVC et la désespérance d'être face à un illuminé. Mais je savais très exactement ce que j'allais dire. Mot pour mot. Je savais la *suite*, et je connaissais ma *fonction*... prof d'info $(n-1)$. J'avais devant moi la vertigineuse et singulière expérience d'un éblouissant déjà-vu conçu et maintenu par une lignée d'initiés. J'avais soudain dans les mains le script et le scénario, de plusieurs instants clefs du passé et, je l'espérais, du futur.

> — « *Le mot récurrence vient du latin* recurrens » commençais-je irrésistiblement « *qui signifie « qui revient en arrière ». Si vous avez cinq minutes, je vous raconte une histoire qui devrait vous faire comprendre les notions de récursivité et de récurrence ».*
> — « *De toute façon maintenant il doit y avoir la queue à la cantine* » lâcha-t-elle comme une plainte résignée.

Une motivation aussi franche requérait une démonstration efficace et, surtout, rapide. Mais je n'avais rien d'autre à faire qu'un petit appel de mémoire sur un problème plus simple.

* * *

C'est une histoire vraie qui commence il y a maintenant quelques années. A l'époque j'étais élève au lycée. Mon professeur de mathématiques et d'informatique, M. Suillereaux, était un jovial barbu ventripotent passionné de ces deux matières,

qui à l'époque, étaient administrativement liées. Il était contagieux dans son enthousiasme, et c'est sans doute grâce à lui que je me suis passionné pour ces matières.

La récursivité et la récurrence sont des notions qui, je l'accorde, ne sont pas toujours évidentes, et à l'époque j'avais moi-même du mal à saisir leur fonctionnement. Le cours avait été très théorique, expliquant qu'il fallait définir le mécanisme de récurrence et la séquence d'initialisation c'est à dire un événement vérifié qui permettait ensuite de dérouler nos conclusions à l'infini. Je connaissais bien ma théorie, mais j'avais du mal à l'assimiler. J'avais l'information mais je n'arrivais pas à en prendre connaissance. Je pouvais appliquer bêtement mais je passais à côté de la portée réelle de ce savoir et cela me laissait une impression d'occasion manquée dont je voulais me défaire.

Alors que j'étais en salle d'étude, je vis M. Suillereaux qui discutait avec le professeur de physique, Mme Boirier. J'attendis patiemment qu'elle parte après s'être une nouvelle fois moquée de la capacité presque artistique avec laquelle il recouvrait son pantalon de craie. Moqueur, je l'avais été moi aussi à l'époque, mais aujourd'hui je compatirais volontiers. J'interpellai mon professeur :

> — « *M. Suillereaux ! Excusez-moi, je peux vous poser une question à propos de la leçon de mardi dernier ?*»
> — ... (à l'époque c'était déjà la bonne réponse)
> — « *Voilà, j'ai du mal à me représenter les phénomènes de récursivité et de récurrence.* » me lançais-je « *Je vois bien la théorie mais j'aurais aimé un exemple concret. Comment peut fonctionner quelque chose d'infini dont la définition se regarde le nombril ?* »

Nous étions près des fauteuils à côté du distributeur de café dans lequel il venait de glisser une pièce, juste avant que je l'interrompe dans cette routine. Alors, il s'était assis avec sur les lèvres un sourire de satisfaction que j'eus le tort de prendre pour un compliment, et un regard trahissant un profond plaisir. Il était tellement perdu dans ses pensées que j'eus la conviction d'être le seul de nous deux à avoir remarqué que la boisson qu'il avait machinalement commandée et finalement laissée sur le distributeur ne sentait pas le café mais le potage de tomates. S'il n'était pas rare de le voir perdu dans ses pensées et commettre de telles inattentions, là il était littéralement et soudainement complètement absent. Au bout d'une longue minute, et devant mon air interrogateur, il avait commencé ainsi :

> — « *Le mot récurrence vient du latin* recurrens *qui signifie « qui revient en arrière ». Si vous avez cinq minutes, je vous raconte une histoire qui devrait vous faire comprendre la notion de récurrence.* »
> — « *De toute façon j'ai loupé le bus de 18h.* » avais-je intelligemment répondu.

* * *

C'est une histoire vraie qui commence il y a maintenant quelques années. A l'époque j'étais élève au lycée. Mon professeur de mathématiques, Mme Sélanti était une dame sèche et droite. Mais la passion pour sa matière faisait briller ses yeux et plisser ses paupières à chaque nouveau résultat qu'elle présentait. C'est sans doute grâce à elle que je me suis passionné pour cette matière.

La notion de récurrence est difficile à appréhender ; j'avais saisi son fonctionnement théorique avec cette condition initiale et le mécanisme de répétition d'un rang à l'autre permettant de dérouler nos conclusions à l'infini mais j'avais besoin de... de l'éprouver. J'attendis patiemment qu'elle sorte le soir de la salle des professeurs.

> — « *Mme Sélanti ?!* » l'interpelai-je « *Je vous prie de m'excuser, mais j'ai beaucoup de difficultés à appréhender le concept de récurrence. Comment peut-on accepter une définition qui se mort la queue* ad vitam æternam *?* »

Avec un de ses rares sourires et sous un regard encore plus pétillant qu'à l'ordinaire, elle m'avait invité à prendre un thé à une terrasse proche avec un engouement que je ne lui avais encore jamais vu. Je me souviens m'être demandé d'où venait cette énergie infinie qui semblait à cet instant l'animer. Devant mon attitude empruntée et mon expression décontenancée elle était finalement redescendue des hautes sphères de ses pensées et avait commencé notre conversation ainsi :

> — « *Le mot récurrence vient du latin* recurrens *qui signifie « qui revient en arrière ». Si vous avez cinq minutes, je vous raconte une histoire qui devrait vous faire comprendre la notion de récurrence.* »
> — « *Mon grand frère m'a prêté sa bicyclette à dynamo je peux rentrer un peu plus tard.* » avais-je confirmé dans un élan de motivation.

* * *

C'est une histoire vraie qui commence il y a maintenant quelques années. A l'époque j'étais élève à l'internat. Mon professeur de mathématiques, M. Polchau, était homme dont la petite taille ne lui permettait pas de remplir la partie haute du tableau, mais il était toujours transporté par ses exposés et explications, au point de ne pas réaliser que ses énormes favoris essuyaient la craie du tableau lorsqu'il se déchaînait sur l'espace noir qui lui était accessible. Son excitation était contagieuse, et c'est sans doute grâce à lui que je me suis passionnée pour cette matière. En matière de pédagogie, je peux dire, pardonnez-moi pour ce trait facile, qu'il avait de la suite dans les idées. Ce jour-là, il avait redoublé d'efforts pour se blanchir les tempes et nous avions tous senti que les notions qu'il nous avait exposées revêtaient pour lui une élégance intellectuelle qui le ravissait. Ce fascinant professeur, qui semblait monté sur ressorts, nous avait notamment proposé deux façons de prouver la valeur de la somme des n premiers entiers : une monstration, c'est à dire une démonstration purement visuelle ou preuve sans mots, disposant deux fois la suite tête-bêche ; et

une démonstration par récurrence qui s'était révélée être un prétexte pour introduire le concept de récursivité.

Le bois des marches, des pupitres et des sièges de l'amphithéâtre craquait encore du passage de mes co-promotionnaires lorsqu'à la fin du cours je m'approchais de l'estrade.

> — « Professeur Polchau ?! » l'interpelais-je, « Je voudrais vous poser une question sur la deuxième partie du cours. »
> — ...

Sur un tableau hissé par une poulie en haut du mur arrière de l'estrade, nous dominait encore la monstration de la somme des entiers. Je la pointais avant de reprendre :

$$\begin{array}{ccccccccc} 1 & + & 2 & + & \ldots & + & (n-1) & + & n \\ n & + & (n-1) & + & \ldots & + & 2 & + & 1 \\ \hline (n+1) & + & (n+1) & + & \ldots & + & (n+1) & + & (n+1) \end{array}$$

> — « J'ai bien compris cette première preuve que vous nous avez donnée, cette monstration, mais pour la deuxième version, vous vous appuyez sur la notion de raisonnement par récurrence, c'est à dire un raisonnement qui étend à tous les termes d'une série, une relation vérifiée pour des termes initiaux. »
> — ...(je n'obtins à ce point qu'un sourire sous un regard espiègle)
> — « Intuitivement, je comprends que ce qui est récurent est ce qui revient périodiquement. » repris-je en essayant de provoquer sa loquacité légendaire. « Mais, dans notre cas, vous dites qu'est récurent ce qui revient en arrière vers son point de départ en utilisant à chaque étape une fonction d'un nombre déterminé d'étapes précédentes. » et je reformulais pour mieux souligner mon blocage « Ce principe semble à la fois remonter à la source mais aussi se propager à l'infini et je dois dire qu'à ce point, je suis perdu. J'ai du mal à saisir ce qui apparaît comme une construction spontanée se propageant instantanément vers le passé et le futur. ».

A mon grand étonnement, son bouillonnement coutumier l'avait quitté, il me regardait, inhabituellement apaisé, sans me voir, comme quelqu'un qui a trouvé dans l'instant présent une porte vers un souvenir et en avait franchi le seuil. J'étais embarrassée par le silence qui s'installait alors qu'il semblait s'être évadé au-delà de ce qui l'entourait, et en toute indifférence pour sa réalité, qu'il savourait ostensiblement un moment suspendu au mépris de mon attente. Je me raclai la gorge et obtins enfin une réaction.

> — « *Le mot récurrence vient du latin* recurrens *qui signifie « qui revient en arrière ». Si vous avez cinq minutes, je vous raconte une histoire qui devrait vous faire comprendre la notion de récurrence.* »
> — « *Je suis à l'internat. J'ai tout mon temps.* » l'avais-je rassuré.

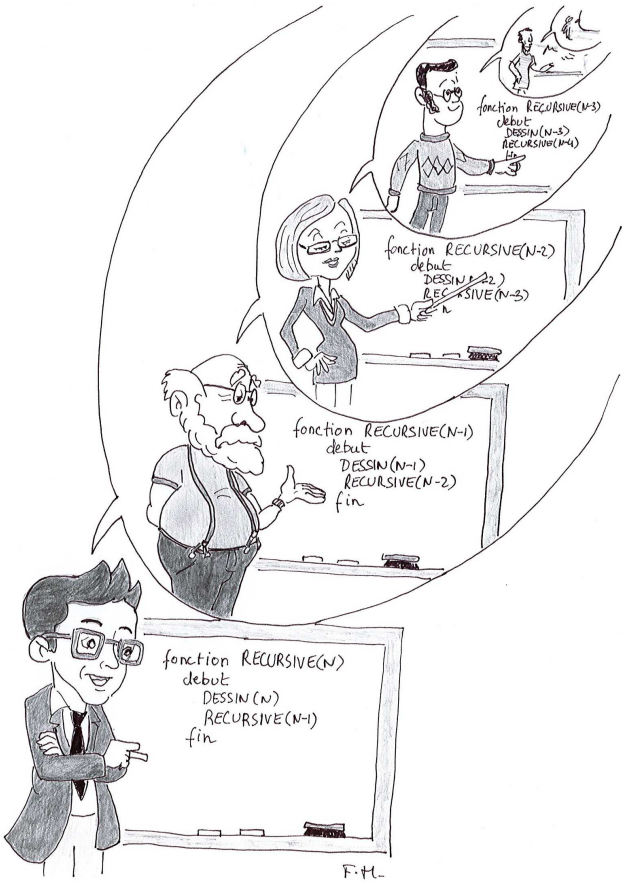
* * *

(…)

Remontée

* * *

Et ainsi de *suite*. Au terme où nous en sommes, vous devriez avoir compris. Assis dans ce rocking chair, le trône neuf d'un professeur d'informatique jeune retraité, ce qui m'amuse le plus dans l'histoire de ce souvenir, c'est que, pour une fois, ce n'est pas l'histoire en elle-même qui nous donne la clef. Ce n'est pas tant le contenu de l'histoire qui nous parle que sa forme qui nous explique. A une époque où tout peut se dire et n'importe comment, je trouverais juste de nous rappeler dans nos messages aux autres que nous devrions plus souvent mettre les formes si nous voulons qu'ils y donnent *suite*.



Être récursif, c'est en particulier pouvoir être, en théorie, répété un nombre infini de fois. Nous n'en aurons pas l'occasion ici, mais ce n'est pas grave puisque nous en connaissons l'essentiel : le terme courant et le principe. Nous pouvons continuer.

Qui était le premier à avoir de la suite dans les idées ? Qui était le hors série ? Zénon d'Élée ? Archimède ? Héron d'Alexandrie ? Pascal ? Roberval ? Bernoulli ? Newton ? Moivre ? Stirling ? Wallis ? Lagrange ? Ou Leonardo Pisano, plus connu sous le pseudonyme de Fibonacci ? Ou n'étaient-ils comme moi à leur époque que le terme suivant ? Peu importe.

Je prends ma retraite. Je viens à terme. Je ne suis plus qu'un des termes précédents et je ne verrai pas la *suite*.

Toutefois, dans la corbeille en osier à côté de mes pieds où s'entassent les cartes de vœux collectées par mes collègues, il y en a une qui vient de quelque part en France et qui me touche beaucoup. Je l'ai reposée ouverte sur le dessus du tas et à chaque fois que le dossier balance vers l'avant je vois le petit mot manuscrit qu'elle porte :

— « *Je prends la suite.* Serial teacher. *Stéphanie.*»

Hommage à Gérard Comyn

Pierre Paradinas [1], Max Dauchet [2]



Gérard Comyn est décédé, lundi 31 janvier 2022, à l'âge de 76 ans. Nous sommes nombreux à avoir croisé Gérard Comyn... Il a légué à tous un peu de sa boussole qu'était le bien commun. Informaticien, chercheur, professeur, directeur de thèse, collègue, il est de ceux avec qui on devient vite proche et avec qui les échanges sont riches et généreux. Débordant de vie, il a marqué par des liens forts toutes celles et ceux qui l'ont connu.

Gérard Comyn a obtenu sa thèse de troisième cycle [3] en 1970, puis sa thèse d'État [4] en informatique en 1982 à l'université de Lille. Il a aussi été chercheur associé au centre d'IBM Paris, puis est parti diriger le centre d'informatique de l'université de Cumana au Venezuela, ce qui valut à Lille d'accueillir plusieurs chercheurs de ce pays.

Jeune maître de conférences en informatique, Gérard Comyn participa avec enthousiasme au développement de ce domaine à l'IUT de Lille1 nouvellement créé. Puis, il fut professeur au LIFL de 1972 à 1989, qui est maintenant la partie informatique du laboratoire CRISTAL.

1. Professeur au Conservatoire national des arts et métiers (CNAM).
2. Professeur des universités, université de Lille.
3. « Simulation de processus biologiques sur ordinateur » dirigée par J.-C. Pages, https://pepite.univ-lille.fr/ori-oai-search/notice/view/univ-lille-19293.
4. « Objet infinis calculables » sous la direction de M. Nivat, https://pepite.univ-lille.fr/ori-oai-search/notice/view/univ-lille-21781.

DOI:10.48556/SIF.1024.20.139 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 139–140

Gérard Comyn pris aussi des responsabilités nationales, en étant l'un des membres de la mission informatique fondamentale et programmation [5] qui publia son rapport en 1983, puis en présidant SPECIF [6] en 1989.

Ses recherches portaient sur les fondements de l'informatique ainsi que leurs applications dans le domaine de la santé.

À la fin des années 80, il œuvrera pour la mise en place d'un projet commun à plusieurs laboratoires d'informatique de Lille (Lifl, Cerim, Trigone) financé par le FEDER, qui permettra l'émergence d'une équipe mixte RD2P [7] impliquant l'industriel de la carte à puce Gemplus.

Il fut nommé en 1989 directeur de l'*European Computer Research Center* (ECRC) fondé par Bull, ICL et Siemens, qui réunissait des chercheurs de toute l'Europe. Il faisait preuve d'une aisance, rare à l'époque pour un français, à passer dans une conversation de l'anglais à l'espagnol ou à l'allemand sans rien perdre de sa chaleur humaine.

Après cette expérience industrielle, il rejoindra la Commission européenne, où il eut en charge au sein de l'unité *Information and Communication Technologies* (ICT) de l'organisation des programmes de recherche dans le numérique, puis les activités relatives à la e-santé.

Durant sa retraite, il se consacre à étendre à l'Europe les activités du réseau associatif national de compétences en télésanté CATEL [8] dont il fut vice-président.

Il a promu avec passion et générosité tout au long de sa vie les valeurs humaines et scientifiques, convaincu qu'elles portent le progrès des connaissances au service du mieux vivre.

5. https://epi.asso.fr/blocnote/Rapport-Nivat-Berry-1983.pdf.

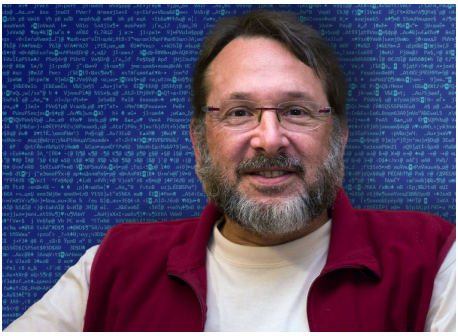
6. SPECIF qui deviendra SPECIF-Campus lors de la création de la SIF en 2012.

7. https://www.cilac.com/sites/default/files/pdf_complet/pi_ndeg73_informatique.pdf, pages 70-71.

8. https://www.catel-esante.fr/.

Hommage à Gilles Muller

Sonia Ben Mokhtar, David Bromberg, Julia Lawall, Laurent Réveillère et Romain Rouvoy



Le 17 novembre 2021, nous avons été bouleversés par l'annonce du décès de notre ami et collègue Gilles Muller, âgé de 59 ans. Il s'est éteint à Rennes après trois années de lutte contre le cancer. Gilles Muller a commencé ses études à Paris 6 en 1980, avant d'intégrer l'INSA de Rennes en 1982. Ayant acquis un amour profond de la recherche, il entame en 1985 un doctorat à l'IRISA dans l'équipe LSP, sous la direction de Jean-Pierre et Michel Banâtre. Depuis sa thèse, portant sur la conception d'une mémoire stable, et tout au long de sa carrière, le fil conducteur de ses recherches a été de rendre les systèmes d'exploitation pleinement fonctionnels et efficaces tout en offrant des garanties de correction.

Il intègre en 1996 l'équipe Compose à Rennes où il explore avec Charles Consel jusqu'en 2002 l'utilisation de langages dédiés pour l'implémentation de pilotes de périphériques. Ces travaux ont donné lieu à Devil [3], un langage dédié offrant des abstractions de périphériques et permettant de vérifier les propriétés de cohérence. Devil a été publié à OSDI en 2000 et a jeté les bases des recherches menées par d'autres sur la connexion des pilotes et des spécifications matérielles, la formalisation du comportement des pilotes et la synthèse des pilotes. La fiabilité des pilotes reste aujourd'hui un domaine de recherche actif.

DOI:10.48556/SIF.1024.20.141 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 141–143

À partir de 2004, dans l'équipe Obasco à Nantes, il a proposé Coccinelle, une solution basée sur des règles de réécriture pour documenter et automatiser les transformations requises pour la maintenance des pilotes sur de multiples versions de Linux. Coccinelle [1, 4] permet de décrire les transformations du code source du noyau Linux lui-même et est maintenant largement adopté dans le développement de celui-ci.

En 2009, Gilles Muller rejoint Inria Paris où il créera en 2015 l'équipe Whisper. Il explore alors l'utilisation de langages dédiés pour connecter les systèmes d'exploitation et les assistants de preuve. Le langage dédié élève le niveau d'abstraction, rendant le code accessible à la preuve, et peut être compilé en un module Linux efficace. Un premier résultat notable est la conception du langage Ipanema [2] qui cible les ordonnanceurs de processus.

En juillet 2021, il a été le premier lauréat du prix chercheur senior décerné par le GDR Réseaux et systèmes distribués (RSD) et le chapitre français de l'ACM SIGOPS, en reconnaissance des nombreuses contributions scientifiques qu'il a développées au cours d'une carrière très riche. Ses derniers travaux portaient sur l'amélioration des performances des applications multicœurs par la réduction du nombre de cœurs utilisés afin d'en augmenter la fréquence d'exécution. Il a participé à l'exploration de cette idée originale jusqu'au bout et a soumis un article à la conférence EuroSys en octobre 2021 qui sera par la suite accepté et présenté en avril 2022.

Les réalisations de Gilles Muller en matière de recherche ont été accompagnées d'un engagement continu dans le renforcement des communautés françaises et européennes en système. Il a notamment a été président du comité de programme d'ACM EuroSys'10 (Paris) et de IEEE DSN'18 (Luxembourg), et a fait partie des comités de pilotage d'EuroSys et de PLOS. Il est un des pères fondateurs de l'ASF, le chapitre français d'ACM SIGOPS. Il a toujours été très fortement impliqué dans les actions du GDR et l'animation de la conférence CFSE, devenue depuis Compas, dont il a été président du comité de pilotage jusqu'au dernier moment. Gilles était un chercheur remarquable et un homme exceptionnel, toujours disponible et prêt à s'engager pour la communauté scientifique. Il n'était jamais avare de son temps pour soutenir les jeunes chercheurs et promouvoir la recherche en systèmes en France et en Europe.

De manière moins formelle, il a collaboré avec de nombreux jeunes chercheurs à travers la France et a encadré de nombreuses thèses en système. Il a constamment encouragé les intérêts de recherche de ces collègues, plutôt que de simplement imposer ses propres objectifs et motivations. Nous garderons le souvenir d'un chercheur passionné, d'un ami sincère et d'un bon vivant, toujours souriant.

Références

[1] Julia Lawall and Gilles Muller. Coccinelle : 10 years of automated evolution in the Linux kernel. In *2018 USENIX Annual Technical Conference (USENIX ATC 18)*, pages 601–614, 2018.

[2] Baptiste Lepers, Redha Gouicem, Damien Carver, Jean-Pierre Lozi, Nicolas Palix, Maria-Virginia Aponte, Willy Zwaenepoel, Julien Sopena, Julia Lawall, and Gilles Muller. Provable multicore schedulers with Ipanema : application to work conservation. In *Proceedings of the Fifteenth European Conference on Computer Systems*, pages 1–16, 2020.

[3] Fabrice Mérillon, Laurent Réveillere, Charles Consel, Renaud Marlet, and Gilles Muller. Devil : An *IDL* for hardware programming. In *Fourth Symposium on Operating Systems Design and Implementation (OSDI 2000)*, 2000.

[4] Yoann Padioleau, Julia Lawall, René Rydhof Hansen, and Gilles Muller. Documenting and automating collateral evolutions in Linux device drivers. In *EuroSys*, pages 247–260. ACM, 2008.

Authentification biométrique et respect de la vie privée : points de vue et limitations

Estelle Cherrier-Pawlowski [1]

Le respect de la vie privée dans les systèmes d'authentification est intrinsèquement lié à la sécurité des données et à l'utilisabilité du système. Le respect de la vie privée englobe à la fois la protection des données personnelles, qui peut être reliée à leur sécurité, via la garantie de leur confidentialité et leur intégrité notamment. D'autre part, si les individus sont confiants dans le niveau (perçu) de protection de leurs données par un système (d'authentification biométrique), cette confiance contribue indiscutablement à l'utilisabilité de ce système.

Cet article propose d'appréhender la notion de vie privée selon différents points de vue : historique, éthique, informatique, réglementaire, avec un focus sur les données biométriques, données personnelles particulièrement sensibles. Une discussion sur les limitations des schémas actuels d'authentification biométrique protégeant la vie privée est proposée à la fin de l'article.

Différents points de vue sur la notion de vie privée

On peut distinguer deux types d'approches pour définir la protection de la vie privée : une première, qui tente de « protéger l'identité de l'utilisateur », et une seconde qui se concentre sur la « protection des données de l'utilisateur ». Dans la première approche, la protection de la vie privée est obtenue en empêchant un adversaire de faire le lien entre les données correspondant à un individu et son identité. Dans le

1. Maîtresse de conférences, HDR, laboratoire GREYC, Ensicaen.

DOI:10.48556/SIF.1024.20.145 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 145–160

second cas, l'identité du propriétaire des données est supposée être connue de l'adversaire. La protection des données est alors obtenue en modifiant les informations divulguées à l'adversaire, en les transformant, par exemple en les remplaçant par une valeur perturbée, en augmentant leur granularité, en y ajoutant des résultats fictifs, etc. On trouve des propositions pour les deux approches dans le cadre de la protection des données de localisation, particulièrement sensibles, dans la thèse de Nicolás E. Bordenabe [1].

La notion même de vie privée possède de nombreuses acceptions, qui toutes dépendent d'un contexte historique, social, culturel, légal. On distingue plusieurs types de vie privée. Chaque type est centré sur un aspect particulier : sur le corps (*bodily privacy*, correspondant à l'intégrité physique), sur les lieux de vie (*locational privacy*, correspondant à la protection du foyer, de la maison), sur les relations interpersonnelles (*relational privacy*, correspondant à la protection de la vie de famille, de la vie sociale) et enfin sur les données (*informational privacy*, correspondant à la protection des données personnelles ainsi qu'au secret des correspondances). Plus de détails sur ces différentes catégories peuvent être retrouvés dans le livre de B. Roessler [21].

La part de la vie privée impliquée dans l'authentification biométrique est par nature en lien avec sa dimension informationnelle. Une typologie plus approfondie de la vie privée au sens de la protection des données personnelles est illustrée par la figure 1, extraite de [15].

De façon générale, un défi se pose lorsqu'on réfléchit à la question : comment mesurer le respect de la vie privée dans un système ? L'article de Wagner et Eckoff [23] propose un tour d'horizon complet des différentes métriques pour évaluer la protection de la vie privée. Ces métriques sont classées selon quatre caractéristiques : les modèles de l'attaquant, les sources de données, les entrées et les sorties du système.

La vie privée est souvent définie par opposition à la vie publique. L'étude de cette notion déborde largement du cadre législatif. En effet elle intéresse de nombreux domaines, comme la sociologie, la médecine ou l'histoire, en passant par l'économie, l'éthique, et l'informatique bien sûr. Cette diversité de traitements rend quasi impossible l'établissement d'une définition universelle. On peut noter toutefois deux termes qui semblent constituer un socle commun, à savoir l'*accès* et le *contrôle*, comme on le voit sur la figure 1.

Dans le cadre de cet article, je vais m'attarder sur plusieurs éclairages reflétant la pluralité des définitions de la vie privée : l'évolution historique du concept de vie privée, des considérations éthiques, la vision informatique, puis réglementaire. Je m'intéresserai enfin aux données biométriques.

Point de vue historique

Cette section retrace en quelques dates et faits marquants les variations de la notion de vie privée au cours des siècles.

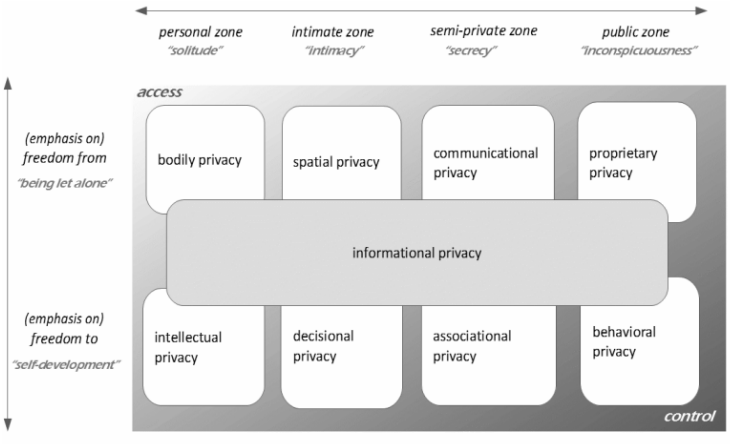


FIGURE 1. Typologie de la vie privée informationnelle [15]

Avant 1500, la notion même d'individualité n'existe pas. Seule la distinction entre la sphère familiale (donc privée, *oikos* en grec) et la sphère publique (principalement les activités politiques, *polis* en grec), développée par Aristote au IV^{e} siècle av. JC, existe.

A partir du XVI^{e} siècle, jusqu'à la Révolution française, le respect de la vie privée recouvre l'acception suivante : autonomie personnelle et individualité. L'émergence d'une classe moyenne ainsi que l'invention de l'imprimerie induisent le développement du travail intellectuel ainsi que la nécessité de disposer d'un endroit calme, privé. Même si la notion de vie privée revêt une connotation négative chez Shakespeare, en raison de son lien avec la solitude, ou encore avec les conspirations politiques, le philosophe anglais John Locke (1632-1704) — un des pères de l'*État de droit* — met à jour le but légitime des individus à protéger leur vie privée, leur liberté et leur propriété.

Au XIX^{e} siècle, les nouvelles technologies de l'époque sont en plein essor : courrier postal, télégraphe, téléphone, enregistrements audio, photographie, etc. On voit alors émerger deux visions de la notion de vie privée : (i) celle de l'état de surveillance, omniscient ; (ii) celle du citoyen, en tant qu'idéal et aspiration de chacun. La première menace directement la vie privée des citoyens, dans le sens où la discipline règne, avec comme point d'orgue le panoptique de Bentham. La seconde vision, celle des citoyens, fait suite aux différentes Révolutions (française, américaine) et Déclarations des droits (*Bill of Rights* en 1689 en Angleterre et Déclaration des droits de l'homme et du citoyen en France en 1789).

Cette tendance se confirme au XX^{e} siècle : en réaction aux deux guerres Mondiales et à la fin du colonialisme, les libertés et les droits fondamentaux sont réaffirmés. La Déclaration universelle des droits de l'homme est ainsi proclamée par l'Assemblée générale des Nations unies le 10 décembre 1948. Elle revêt uniquement une valeur déclarative. L'article 12 concerne la vie privée Un traité international est ensuite signé le 4 novembre 1950 par les états membres du Conseil de l'europe et entre en vigueur le 3 septembre 1953 : la Convention européenne des droits de l'homme (CEDH, aussi appelée Convention de sauvegarde des droits de l'homme et des libertés fondamentales). Afin d'assurer le respect des engagements de la CEDH, une juridiction internationale est instituée en 1959 à Strasbourg : la Cour européenne des droits de l'homme.

A partir des années 1970, la vie privée s'entend à l'ère des ordinateurs comme la protection, la confidentialité des informations (*information privacy*). Depuis les attaques du 11 septembre 2001 et la lutte contre le terrorisme qui s'est renforcée ensuite, la vie privée est un droit fondamental que les états s'autorisent à rogner sous prétexte de renforcer la sécurité des citoyens. Contre l'argument « *personne ne peut être contre les mesures de sécurité s'il n'a rien à cacher, rien à se reprocher* », Edward Snowden, lanceur d'alerte américain rétorque [2] : « *Dire que votre droit à la vie privée importe peu car vous n'avez rien à cacher revient à dire que votre liberté d'expression importe peu, car vous n'avez rien à dire. Car même si vous n'utilisez pas vos droits aujourd'hui, d'autres en ont besoin. Cela revient à dire : les autres ne m'intéressent pas* ». Il apparaît donc que la notion même de vie privée est un concept fluctuant au cours de l'histoire, en constante évolution donc, qui ne peut être étudié hors d'un contexte social, politique, économique et international.

Point de vue éthique

Comme mentionné dans le livre édité par Bart van der Sloot et Aviva de Groot [6, chapitre 5], un point de vue éthique sur la vie privée permet de répondre aux questions suivantes : quelle est la valeur de la vie privée ? Quelles normes concernant la vie privée doivent être respectées par les individus, la société, l'état ? Même si certains chercheurs renoncent à figer une définition précise — forcément réductrice — de la notion de vie privée, un grand nombre de définitions incluent les termes *accès* et *contrôle*.

Tout d'abord on peut citer la définition de Reiman [20] : « *privacy is the condition in which others are deprived of access to you* ». Le terme *accès* signifie soit un accès physique à une personne, soit un accès à des informations sur cette personne. Le point de vue éthique s'intéresse alors (i) à la manière dont cet accès est obtenu ; (ii) à quoi exactement on a accès. Cependant Fried [10] a montré que l'*accès*, seul, ne suffit pas. Il faut lui adjoindre le *contrôle*, ceci pour éviter le cas absurde d'un homme isolé sur une île déserte : effectivement, personne n'a accès à lui, mais la

2. *Nothing to hide*, documentaire produit par Mihaela Gladovic et Marc Meillassoux, 2017.

notion de vie privée n'a aucun sens dans ce contexte. La définition de Fried relie les deux notions : le respect de la vie privée consiste à contrôler l'accès à soi. En fait, l'accès à une personne ou à des informations personnelles n'est pas à rejeter en bloc : l'important est de pouvoir contrôler cet accès.

On peut remarquer que, d'une part, ces définitions se réfèrent à l'article *The Right to Privacy* de Warren et Brandeis [25] ; d'autre part, elles mettent en exergue la question éthique suivante : est-ce que les protections légales qui existent sont suffisantes pour protéger les droits individuels, en particulier le contrôle exercé par chaque individu sur sa vie privée ? Cette question est d'autant plus cruciale que les technologies actuelles tendent à défier les normes sociales. Les travaux de Beate Roessler [6, p. 137] proposent une réponse selon trois dimensions de la vie privée :

— la dimension locale : comment chacun contrôle l'accès à ses espaces physiques (par exemple, la clé sur la porte d'entrée de sa maison) ;
— la dimension informationnelle (le terme *informations* renvoie souvent aux *données*) : contrôle sur ce que les autres peuvent savoir de nous ;
— la dimension décisionnelle : contrôle sur l'accès (symbolique) à nos décisions personnelles, nos valeurs, nos objectifs, les raisons de nos décisions, etc.

Selon B. Roessler [21], ces trois dimensions de la vie privée doivent actuellement faire face à des défis.

Défis de la dimension locale. Les objets connectés sont par définition connectés à Internet, sphère publique et totalement ouverte. Simultanément, ils envahissent notre espace privé, ce qui n'est pas inquiétant en soi. Ce qui est potentiellement dangereux, c'est la collecte, le stockage, l'analyse, etc., à l'intérieur même de nos maisons, de nos données privées par des assistants personnels, des compteurs électriques ou autres radiateurs connectés, alors que notre espace physique privé était à peu près impénétrable jusque-là. Ces objets dits *intelligents* représentent également une menace potentielle pour les autres dimensions informationnelle et décisionnelle de notre vie privée.

Défis de la dimension informationnelle. Comme on vient de le voir, la collecte de données dans des domaines d'activité variés est de plus en plus importante. Couplée à des algorithmes d'apprentissage et de fouille de données toujours plus efficaces, cette collecte mène à la création de profils qui peuvent être utilisés dans différents contextes ; ainsi nos identités (numériques) ne peuvent plus être cloisonnées. La persistance des traces informatiques constitue une autre source d'inquiétude : dès qu'une donnée est créée, elle peut aussitôt être partagée très facilement et exploitée dans un contexte inconnu, dans un but inconnu. Dans son best-seller *L'Âge du capitalisme de surveillance* [26], Shoshana Zuboff retrace l'histoire de l'*économie de surveillance* : celle-ci repose sur un principe de subordination et de hiérarchie. Elle met en évidence le projet consistant à extraire une plus-value de nos agissements sur

Internet. L'auteur dénonce que Google a découvert que nous avions moins de valeur que les pronostics que d'autres font de nos agissements : « *L'industrie numérique prospère grâce à un principe presque enfantin : extraire les données personnelles et vendre aux annonceurs des prédictions sur le comportement des utilisateurs. Mais, pour que les profits croissent, le pronostic doit se changer en certitude. Pour cela, il ne suffit plus de prévoir : il s'agit désormais de modifier à grande échelle les conduites humaines.* »

Défis de la dimension décisionnelle. Pour continuer sur la même thématique, le constat suivant est un doux euphémisme : le commerce de nos données, à travers l'essor du *Big Data*, de l'*IoT*, des réseaux sociaux augmente le risque que ces technologies nous influencent, manipulent nos comportements. En 2014, Facebook a lancé une étude sur la manipulation des émotions, sans le consentement de ses utilisateurs. D'un point de vue éthique, le problème n'est pas le sujet de l'étude, mais les conditions dans lesquelles elle a été réalisée, sans respecter les normes scientifiques et encore moins éthiques, comme le montre l'analyse de Flick dans son article [9]. De la même façon, les média et réseaux sociaux aujourd'hui sont accusés de scléroser la pensée des internautes, en leur suggérant uniquement des contenus qui les intéressent déjà. On se retrouve piégé dans de véritables *bulles de filtres*. La curiosité intellectuelle n'est pas encouragée, impactant de fait la dimension décisionnelle de notre vie privée.

Point de vue informatique (TIC)

L'informatique fait partie des technologies de l'information et de la communication (TIC) en permettant notamment le traitement, le stockage et la communication de données, y compris de données personnelles. La multiplication des collectes, des flux, des traitements de ces données personnelles impacte nécessairement la vie privée des individus. En fait, c'est la relation faite entre des données capturées et l'identité des utilisateurs qui représente un risque pour la vie privée. La vie privée numérique est intrinsèquement liée à la sécurité numérique, dans le sens où elle relève de la confidentialité des données. Avec la prolifération des objets connectés, des applications mobiles, les risques sur la sécurité, et en particulier sur la vie privée, se multiplient. On est donc ramené au sujet plus général de la gestion des risques en sécurité informatique. Cependant, aucune méthode générique d'analyse des risques sur la vie privée n'existe, contrairement aux méthodes d'analyse des risques de sécurité. Il est donc très difficile d'évaluer le niveau de respect de la vie privée propre à un système, comme il est très difficile de comparer deux systèmes en matière de respect de la vie privée. L'article de Wagner et Eckoff [23] le prouve en recensant plus de quatre-vingt métriques disponibles sur le sujet.

Les données des utilisateurs peuvent être dans trois états selon [6, chapitre 5] : (i) les données stockées, ou statiques (i.e. *data at rest*, par exemple dans une archive, ou

une base de données peu consultée) ; (ii) les données utilisées, ou en cours d'utilisation (i.e. *data in use*, par exemple dans une base de données fréquemment modifiées par de multiples utilisateurs, ou des données en cours de traitement) ; (iii) les données en mouvement (i.e. *data in motion*, lorsque les données sont transférées via des réseaux informatiques à l'extérieur du système, ou sont utilisées par des services tiers). Les données ne sont pas soumises aux mêmes risques dans chacun des trois états.

Avec l'essor des ordinateurs personnels, puis le développement d'Internet et aujourd'hui l'essor des *smartphones* et des objets connectés dits *intelligents*, de véritables défis sont posés dans le domaine informatique, par la société ou la technologie, comme l'éthique de l'intelligence artificielle (dans les véhicules autonomes par exemple), l'éthique du *Big Data* (comment éviter les dérives dans les traitements de données massives), le développement de technologies respectueuses de la vie privée (ou *Privacy Enhancing Technologies*, PET).

On peut donc constater que le problème du respect de la vie privée est de plus en plus prégnant dans notre société : les nouvelles technologies représentent un défi toujours renouvelé pour les normes sociales. La question suivante se pose : est-ce que les protections légales existantes sont suffisantes pour protéger les droits des individus ? Le paragraphe suivant va donner quelques éléments de réponse.

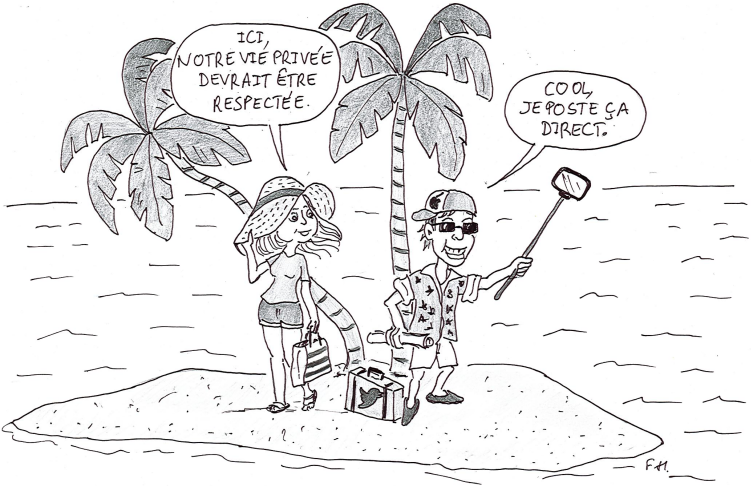
La réglementation sur le respect de la vie privée

Le respect de la vie privée, un droit fondamental

La définition de la vie privée constitue sans doute le plus ancien principe juridique, en tant que séparation entre la sphère publique et la sphère privée. Son étymologie vient du latin *privare* (priver de), et par extension signifie soustraire quelque chose du domaine public. Pendant longtemps, la sphère publique a été gérée par le roi, ou le législateur de façon générale. La sphère privée, quant à elle, était gérée par le *pater familias*, le père de famille, qui régnait sur le foyer, le ménage. Le droit au respect de la vie privée était inclus dans les lois constitutionnelles de certains pays dès le XIII^{e} siècle (en ce qui concerne l'intégrité physique, les lieux privés, les relations privées et le secret des correspondances). Au XX^{e} siècle, le droit à la vie privée figure dans l'article 12 de la Déclaration universelle des droits de l'homme des Nations unies de 1948. Ce droit figure également dans les articles 7 (*« respect for private and family life »*) et 8 (*« protection of personal data »*) de la Convention européenne des droits de l'homme de 1950 [3]. La Charte des droits fondamentaux de l'Union européenne, proclamée en 2000, comporte cinquante-quatre articles consacrant les droits fondamentaux des personnes, y compris les droits sociaux. Elle englobe le droit à la vie privée. Elle est contraignante pour les pays de l'UE dans le sens où elle se voit reconnaître une valeur constitutionnelle par le Traité de Lisbonne

3. https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005.

en 2007. En France, c'est le Code Civil qui reconnaît le droit à la vie privée : selon l'article 9 alinéa 1, « *Toute personne a droit au respect de sa vie privée* ». Comme le droit à l'image ou le droit à l'honneur, le droit à la vie privée fait partie des *Droits de la personnalité* : ce sont des droits inhérents à la personne humaine, qui la protègent en interdisant toute atteinte à ses droits les plus fondamentaux (sa vie, sa dignité, son corps).



Dans la suite de cet article, seuls les aspects informatiques de la vie privée ainsi que la protection des données personnelles seront considérés, qui concernent directement les systèmes biométriques. C'est au niveau européen que la réglementation évolue : tout d'abord, le principe de *Privacy by Design* (PbD) (ou respect de la vie privée dès la conception), puis le règlement général sur la protection des données (RGPD).

Privacy by design et RGPD

Le *Privacy by Design* a été élaboré par Ann Cavoukian (commissaire à l'information et à la protection de la vie privée de l'état d'Ontario) vers la fin des années 1990 au Canada. Ce concept, détaillé dans le document [2], est une réponse à l'automatisation du traitement des données personnelles et à la quantité de plus en plus importante de données manipulées par les entreprises (ne nécessitant souvent pas une intervention humaine). Le *Privacy by Design* est une mesure préventive ayant donc pour but de limiter les risques d'abus et de violation des données : il implique une prise en compte de la protection de la vie privée des utilisateurs avant même la conception d'un système impliquant le traitement de données personnelles.

Le *Privacy by Design* énonce sept principes [4] :

— conception de mesures préventives et proactives ;

4. https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf.

— protection par défaut (*Privacy by default*) ;
— prise en compte des règles sur la protection de la vie privée dans la conception des produits et durant leur utilisation ;
— protection optimale et intégrale ;
— sécurité de bout en bout – protection du cycle de vie ;
— visibilité et transparence ;
— respect de la vie privée des usagers ou des cibles du service.

Le *Privacy by Default* est un principe qui vise tout produit ou service rendu public. Les standards en matière de protection des données personnelles s'appliquent alors par défaut, et ce sans recours à des procédures extérieures permettant cette protection. Le *Privacy by Default* est la garantie d'un niveau maximal de protection des données personnelles.

Les principes énoncés dans le *Privacy by Design* ont largement été repris dans le RGPD : en mai 2018, le règlement général sur la protection des données [8] vient remplacer la directive 95/46/CE du Parlement européen et du conseil du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données. La définition d'une donnée personnelle est clairement posée par la CNIL [5] :

Définition 1 (Donnée à caractère personnel)**.** *Une donnée à caractère personnel est toute information se rapportant à une personne physique identifiée ou identifiable. Mais, parce qu'elle concerne des personnes, celles-ci doivent en conserver la maîtrise.*

Parmi les données personnelles, certaines données sont considérées comme particulièrement sensibles. Le RGPD interdit de recueillir ou d'utiliser ces données, sauf, notamment, si la personne concernée a donné son consentement exprès (démarche active, explicite et de préférence écrite, qui doit être libre, spécifique, et informée). Ces exigences concernent par exemple les données de santé, les opinions politiques, les convictions religieuses, ou encore les données biométriques. Par conséquent, en tant que donnée personnelle sensible, toute donnée biométrique devrait être protégée et ne devrait pas être stockée, ni transiter entre deux éléments d'un système biométrique sous sa forme brute, ni sous une forme qui permet de reconstruire (une partie de) la donnée originale.

La recherche en biométrie est particulièrement concernée par l'article 35 du RGPD, qui introduit la notion d'analyse d'impact relative à la protection des données (AIPD, ou DPIA pour *Data Protection Impact Assessment*). Cette analyse doit être réalisée dès lors qu'un traitement sur des données est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes concernées. Toutefois, il est important de souligner que les données biométriques, dans un contexte de recherche

5. https://www.cnil.fr/fr/definition/donnee-personnelle.

publique, ont un statut spécial : le RGPD prévoit quelques options, offrant des latitudes aux pays membres de l'UE, notamment en ce qui concerne les données de recherche. La France a choisi d'assouplir les contraintes, pour respecter aux mieux les objectifs de recherche. Ainsi, la durée de conservation peut être allongée, la finalité peut évoluer au cours du temps... On trouve une liste de mesures appropriées à mettre en place pour pouvoir collecter des données biométriques dans un cadre de recherche sur le site de la CNIL, dans un document [6] intitulé Présentation du régime juridique applicable aux traitements poursuivant une finalité de recherche scientifique (hors santé) :

— conformément à l'article 5.1-c) du RGPD, le principe de pertinence et de minimisation des données traitées doit être respecté ;
— l'anonymisation des données pourrait être mise en œuvre (le G29 a publié des recommandations à ce sujet en 2014) ; elle est obligatoire lors de la diffusion des résultats, diffusion nécessaire à la présentation des travaux de recherche (cf. l'article 116 du décret n°2019-536 du 29 mai 2019) ;
— la pseudonymisation doit être mise en œuvre toutes les fois où cela s'avérerait pertinent (sachant que les informations permettant de faire le lien entre les données et l'identité des individus doivent être conservées séparément et soumises à des mesures techniques et organisationnelles) ;
— une logique d'accès sécurisé et contrôlé doit être développée, en fonction de la sensibilité des données et des finalités des réutilisations futures ;
— la réalisation d'une analyse d'impact sur la protection des données est quasi systématique.

L'article de Sanchez et al. [22] propose une adaptation de ces mesures dans la pratique. On constate donc que l'adoption du RGPD a permis de clarifier et de structurer le cadre à respecter pour la collecte de données biométriques : le fait même qu'à partir d'une bribe de donnée biométrique, on pourrait être en mesure d'extraire une information nuisible pour la vie privée de l'utilisateur rend la protection de ces données indispensable. C'est précisément l'objet des techniques de biométrie révocable.

Limitations

Même si le législateur a inévitablement un peu de retard sur les risques sur la vie privée dus aux nouveaux usages des *smartphones* et autres objets connectés, on constate que les protections des citoyens évoluent dans le bon sens, que ce soit au niveau du droit français, ou européen, comme le montre la progression du respect de la vie privée dans la hiérarchie des normes.

6. https://www.cnil.fr/sites/default/files/atoms/files/consultation_publique_-_presentation_du_regime_juridique_applicable_aux_traitements_a_des_fins_de_recherche.pdf.

Dans son article, Els Kindt pointe les insuffisances du RGPD en matière de protection des données biométriques [14], notamment à travers ce constat (Art. 4, définition 14 du RGPD[7]) : « *Les données biométriques sont les données à caractère personnel résultant d'un traitement technique spécifique, relatives aux caractéristiques physiques, physiologiques ou comportementales d'une personne physique, qui permettent ou confirment son identification unique* ».

Deux éléments sont donc nécessaires pour qu'une donnée biométrique entre dans ce cadre : le traitement spécifique et l'identification unique. Comme contre-exemple, E. Kindt propose une base de données de photographies d'empreintes ou de visages, simplement collectées et stockées : dans ce cas, ces images ne seraient pas des données biométriques (cf. le considérant 51 du RGPD). Elle poursuit son analyse en étudiant la distinction entre quatre catégories de « données personnelles en lien avec des caractéristiques physiques, physiologiques ou comportementales d'un individu, et qui permettent son identification ou authentification » :

— *les données personnelles ordinaires* (considérées comme non biométriques) : les données simplement stockées, les données de biométrie douce (qui ne permettent pas une identification unique). C'est le régime général du RGPD qui s'applique à ces données (Art. 6) ;

— *les données biométriques* : cette catégorie, qui concerne les données biométriques en général (qui subissent un traitement spécifique et qui permettent ou confirment une identification unique) sont sujettes au même régime que les données personnelles ordinaires ;

— *les données biométriques sensibles* : ce sont les données qui sont traitées dans le seul but d'identifier de façon unique des individus. Dans ce cadre, l'article 9 du RGPD s'applique également : en principe, l'identification biométrique est interdite, sauf exceptions. Il faut souligner qu'il s'agit ici uniquement d'identification. L'authentification pourrait, sous couvert d'une certaine interprétation du RGPD, tomber dans la catégorie précédente (le but n'est pas une identification unique). Cependant, E. Kindt recommande des conditions supplémentaires : l'authentification devrait être transparente pour l'utilisateur, avec des garanties légales et techniques (pas de stockage dans une base centralisée), et le recours obligatoire à une AIPD ;

— *les données biométriques dont le but est d'identifier des individus à grande échelle* : ces données sont sujettes à des contraintes réglementaires additionnelles (Art. 35 et 36 du RGPD).

L'article de Gellert [11] porte sur la notion de risque, telle qu'elle est définie dans le RGPD. Il propose également une réflexion très pertinente sur la distinction entre *Privacy Impact Assessment* (PIA) et *Data Protection Impact Assessment* (DPIA). L'auteur rappelle que la vie privée va bien au-delà des données personnelles. Par

7. https://gdpr-text.com/fr.

conséquent, mener un DPIA serait forcément réducteur par rapport à un PIA. Il s'appuie sur les travaux de R. Clarke [5] : *« A Data Privacy Impact Assessment is a study of the impacts of a project on only the privacy of personal data, whereas a PIA considers all dimensions of privacy »*. Les dimensions de la vie privée auxquelles il fait référence sont : les informations personnelles, l'intimité de la personne, le comportement individuel, les communications personnelles, les pensées et les émotions. L'auteur, R. Gellert, dépasse ce clivage traditionnel et analyse en profondeur la notion de risque présente dans le RGPD. Il admet que le texte du règlement contient la définition de l'événement (à l'origine du risque), les conséquences (du risque) et un certain nombre de critères liés aux facteurs de risque. Il note toutefois qu'il manque un certain nombre d'éléments constitutifs d'un risque : (i) aucune mention n'est faite concernant la vraisemblance (ou probabilité d'occurrence) du risque (seul son niveau de gravité est mentionné), (ii) il y a peu de critères concernant les conséquences des risques : cela signifie que le règlement laisse une totale liberté quant au calcul des « risques relatifs aux droits et libertés des personnes concernées » dans la pratique, et (iii) le RGPD ne préconise aucune méthodologie pour réaliser un DPIA. La dernière phrase de l'article est la suivante :« *Thus, to some extent, the way risk is defined and understood in the GDPR is pretty much irrelevant* ». Els Kindt et Raphaël Gellert sont des juristes, aussi leur point de vue externe au domaine de l'informatique permet une prise de recul nécessaire et salutaire, en élargissant les perspectives.

On trouve également dans la littérature des critiques plus techniques sur les algorithmes de biométrie révocable. Le concept de *biométrie révocable* [19] repose sur une transformation des données biométriques brutes, de telle sorte que les données transformées soient sûres et respectueuses de la vie privée, en accord avec les propriétés suivantes (cf. [16, 13] ainsi que la norme ISO/IEC 24745) : non-inversibilité, performance, diversité, révocabilité.

Dans l'article [18], Nandakumar et Jain soulignent que les propriétés de révocabilité et de non-inversibilité ne sont pas garanties intrinsèquement dans les schémas de biométrie révocable, et qu'il est nécessaire de développer des fonctions à sens unique pour rendre le calcul d'une pré-image plus difficile. D'autres attaques sont possibles visant les techniques à base de salage des données, elles sont détaillées dans l'article de Choudhury et al. [4] : ces attaques exploitent la présence de la donnée biométrique originale, de la clé, de la donnée transformée, ou l'interception de plusieurs données transformées. Dans l'article [18], les auteurs proposent trois raisons principales au fossé existant entre la théorie (notamment en ce qui concerne les propriétés attendues) et la pratique pour les systèmes de biométrie révocable :

— la méthode d'extraction des caractéristiques (*features*) ;
— le compromis entre la non-inversibilité et les performances de reconnaissance ;
— la garantie des propriétés de non-associativité et révocabilité.

Concernant les schémas reposant sur une transformation des données, souvent la non-inversibilité y est estimée de façon empirique à partir de la complexité de calcul de l'attaque par inversion. Les auteurs citent l'exemple de la courbe *Coverage-Effort* proposée par Nagar et al. [17], qui permet de mesurer le nombre de tentatives (ou efforts) requis pour récupérer une fraction (*coverage*) de la donnée originale. Ils soulignent également que cette approche empirique ne permet pas de traiter toutes les stratégies d'inversion des gabarits par des imposteurs. Dong et al. proposent d'exploiter la propriété de préservation des distances (inhérente aux techniques de biométrie révocable par construction) pour définir une classe d'attaques basées sur la similarité [7]. Les auteurs se placent dans l'hypothèse de Kerckhoff : l'attaquant a accès à un gabarit transformé et connaît la fonction de transformation ainsi que ses paramètres. Ils testent leur méthodologie d'attaque sur plusieurs schémas de biométrie révocable. Dans l'article [3], Chen et al. vont plus loin et démontrent, que lorsque la distance entre deux données biométriques non protégées augmente, la distance entre les données protégées correspondantes augmente également, inévitablement. La solution proposée repose sur le choix d'une fonction de transformation plus compliquée, non linéaire. Gomez-Barrero et al. proposent explicitement un cadre pour évaluer la non-associativité des données protégées par les schémas de biométrie révocable [12]. Les auteurs définissent deux fonctions — une globale et une locale — pour mesurer un score de non-associativité entre deux échantillons, selon qu'ils sont appariés ou non.

Aujourd'hui encore, il est quasi impossible de prouver qu'un schéma de biométrie révocable est totalement résistant aux attaques de sécurité et sur la vie privée. L'article de Wang et Baskerville [24], propose une vision divergente, qui critique les bénéfices de l'authentification à deux facteurs (par rapport à une authentification à un facteur), avec une analyse originale risques/bénéfices/coûts. Tous ces éléments laissent donc de nombreuses perspectives de recherche, dans des directions variées.

Conclusion

Pour conclure, voici un extrait d'une interview d'Alain Damasio (écrivain de science-fiction), diffusée sur France Culture en 2019[8] : « *Ce qu'il se passait en milieu fermé dans la prison où on était hypra surveillés, est devenu la norme d'existence aujourd'hui. Ça ne gêne personne au sens où tant qu'on n'a pas le retour de pouvoir sur nos vies, tant que quelqu'un ne s'empare pas de ça pour nous « emmerder », pour nous bloquer, pour nous refuser un job, on ne sent pas qu'on est pris dans cette nasse. Mais le fait est qu'on circule dans des prisons à ciel ouvert en fait.* »

Cette réflexion est également présente dans le livre de S. Zuboff [26], qui tente d'expliquer pourquoi on a « laissé » Google, Facebook, Instagram, etc., aspirer nos

8. https://www.franceculture.fr/societe/le-panoptique-a-lorigine-de-la-societe-de-surveillance.

données et nos comportements : cette révolution numérique est *sans précédent* (formule de l'auteure), par conséquent les utilisateurs n'ont pas de repères, pas de modèles auxquels se raccrocher. On a donc laissé faire au départ, de façon naïve et innocente, jusqu'à ce qu'on ne puisse plus se passer des services proposés, qui simplifient la vie quotidienne, et qu'on ferme les yeux sur le prix à payer.

L'Union européenne a réagi, sans doute un peu tard, en déployant le RGPD pour protéger les données de ses ressortissants. Ce règlement fait des émules, en Californie par exemple. Cependant, les entreprises s'adaptent très vite aux différentes contraintes : les lignes directrices du Conseil européen de protection des données (EDPB) sur le consentement valide publiées en mai 2020 stipulent que les *cookie walls* ne sont pas un moyen valide pour les sites web d'obtenir le consentement des utilisateurs au traitement de leurs données personnelles et à l'utilisation de *cookies*. Pourtant, on constate aujourd'hui que certains sites web ont trouvé la parade en quelques semaines en rendant leur consultation payante si les *cookies* sont refusés par les utilisateurs. Ce comportement présente au moins l'avantage de mettre fin à l'hypocrisie du « tout gratuit » sur internet. Le livre [26] met fin à la croyance classique « si c'est gratuit, c'est toi le produit » : les géants d'internet ne cherchent plus à créer des profils d'utilisateurs, ils recourent aujourd'hui à l'intelligences artificielle pour prédire leurs comportements futurs, et c'est cette prédiction qu'ils revendent à des annonceurs, des services marketing, des publicitaires, etc., qui à leur tour vont essayer d'influencer les comportements des utilisateurs.

Il est donc peut-être temps de renverser le modèle et de replacer l'utilisateur au cœur de la collecte de ses données, en appliquant la transparence, la minimisation, le consentement éclairé, imposés par l'application du RGPD. On peut également aller plus loin en redonnant à l'utilisateur une souveraineté sur ces aspects, une compréhension des enjeux et un réel pouvoir de décision, par exemple à travers le développement d'une application de collecte de données respectueuse de la vie privée : il s'agit de mon projet de recherche, exposé en détail dans mon manuscrit d'habilitation à diriger des recherches (HDR) [9].

Pour finir, je voudrais citer G. Bernanos, qui écrivait déjà en 1947 [10] :« *L'idée qu'un citoyen, qui n'a jamais eu affaire à la justice de son pays, devrait rester parfaitement libre de dissimuler son identité à qui il lui plaît, pour des motifs dont il est seul juge, ou simplement pour son plaisir, que toute indiscrétion d'un policier sur ce chapitre ne saurait être tolérée sans les raisons les plus graves, cette idée ne vient plus à l'esprit de personne.* »

Références

9. https://tel.archives-ouvertes.fr/tel-03326656.
10. La France contre les robots, Georges Bernanos, 1947.

[1] Nicolás E. Bordenabe. *Measuring Privacy with Distinguishability Metrics :Definitions, Mechanisms and Application to Location Privacy*. PhD thesis, Ecole Polytechnique, 2014.

[2] Ann Cavoukian. Privacy by design, 2011.

[3] Yanzhi Chen, Yan Wo, Renjie Xie, Chudan Wu, and Guoqiang Han. Deep Secure Quantization : On secure biometric hashing against similarity-based attacks. *Signal Processing*, 2019.

[4] Bismita Choudhury, Patrick Then, Biju Issac, Valliappan Raman, and Manas Kumar Haldar. A Survey on Biometrics and Cancelable Biometrics Systems. *International Journal of Image and Graphics*, 2018.

[5] Roger Clarke. An evaluation of privacy impact assessment guidance documents. *International Data Privacy Law*, 2011.

[6] Aviva de Groot and Bart van der Sloot, editors. *The Handbook of Privacy Studies : An Interdisciplinary Introduction*. Amsterdam University Press, 2018.

[7] Xing-Bo Dong, Zhe Jin, A. B. J. Teoh, M. Tistarelli, and K. Wong. On the reliability of cancelable biometrics : Revisit the irreversibility. 2019.

[8] European Parliament. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Available : `http://eur-lex.europa.eu/eli/reg/2016/679/oj`, 2016.

[9] Catherine Flick. Informed consent and the Facebook emotional manipulation study. *Research Ethics*, 2016.

[10] C. Fried. *Privacy : A Moral Analysis*, pages 203–222. Cambridge University Press, 1984.

[11] Raphaël Gellert. Understanding the notion of risk in the General Data Protection Regulation. *Computer Law & Security Review*, 2018.

[12] Marta Gomez-Barrero, Javier Galbally, Christian Rathgeb, and Christoph Busch. General Framework to Evaluate Unlinkability in Biometric Template Protection Systems. *IEEE Transactions on Information Forensics and Security*, 2018.

[13] Anil K. Jain, Karthik Nandakumar, and Abhishek Nagar. *Fingerprint Template Protection : From Theory to Practice*. Springer London, 2013.

[14] Els Kindt. Having yes, using no ? about the new legal regime for biometric data. *Computer Law & Security Review*, 34, 12 2017.

[15] Bert-Jaap Koops, Bryce Newell, Tjerk Timan, Ivan Skorvanek, Tom Chokrevski, and Maša Galič. A typology of privacy. *University of Pennsylvania Journal of International Law*, 2016.

[16] Davide Maltoni, Dario Maio, Anil K. Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition*. Springer London, 2009.

[17] Abhishek Nagar, Karthik Nandakumar, and Anil K. Jain. Biometric template transformation : a security analysis. In *Electronic Imaging*, 2010.

[18] Karthik Nandakumar and Anil K. Jain. Biometric Template Protection : Bridging the performance gap between theory and practice. *IEEE Signal Processing Magazine*, 2015.

[19] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 2001.

[20] J. Reimann. Driving to the panopticon : A philosophical exploration of the risks to privacy posed by the highway technology of the future. *Santa Clara High Technology Law Journal*, 11(1) :27–44, 1995.

[21] Beate Roessler. *The Value of Privacy*. Cambridge : Polity Press, 2005.

[22] Raul Sanchez-Reillo, Ines Ortega-Fernandez, Wendy Ponce-Hernandez, and Helga C. Quiros-Sandova. How to implement eu data protection regulation for r&d in biometrics. *Computer Standards & Interfaces*, 2019.

[23] Isabel Wagner and David Eckhoff. Technical privacy metrics : A systematic survey. *ACM Computing Surveys*, 2018.

[24] Pengcheng Wang and Richard Baskerville. The case for two-factor authentication- evidence from a systematic literature review. In *Twenty-Third Pacific Asia Conference on Information Systems*, 2019.

[25] Samuel D. Warren and Louis D. Brandeis. The right to privacy. *Harvard Law Review*, 1890.

[26] Shoshana Zuboff. *Age of Surveillance Capitalism : The Fight for a Human Future at the New Frontier of Power*. New York : PublicAffairs, 2019.

Journée « Raconter la science en temps de crise »

Journée Sciences et médias 2022

La pandémie de covid a placé la science au cœur de l'information. Comment les journalistes et les scientifiques ont-ils collaboré ? Comment lutter contre la désinformation ? Quelles leçons tirer de ces deux ans, notamment pour les crises plus longues comme le réchauffement climatique ? Scientifiques et journalistes ont tenté de répondre conjointement.

Quelle joie d'assister à un colloque « en vrai » ! Le 25 janvier 2022 a eu lieu la sixième édition de Sciences et médias, organisée par plusieurs sociétés savantes [1] et l'association des journalistes scientifiques de la presse d'information (AJSPI), ainsi que la BNF. Avec un thème dans l'actualité : raconter la science en temps de crise.

Nous sommes accueillis par Michel Netzer, directeur du département Sciences et techniques à la BNF. Daniel Hennequin, de la SFP, a été frappé par la façon dont la démarche scientifique a été médiatisée. « *Le grand public est rarement confronté à la science en train de se faire, mais plus souvent à des résultats établis, sauf en cas de crise* », rappelle-t-il.

Optimisme

Yves Sciama, journaliste scientifique, vice-président de l'AJSPI, nous apporte un message d'optimisme. Pour la première fois en 25 ans de métier, il observe que la place et le traitement de l'information se sont améliorés pendant cette crise. Même si tout n'est pas rose, « *le journalisme scientifique n'a jamais été aussi fort, aussi*

1. Société française de physique, Société chimique de France, Société française de statistiques, Société informatique de France, Société de mathématiques appliquées et industrielles, Société mathématique de France.

DOI:10.48556/SIF.1024.20.161 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 161–169

visible, et autant reconnu comme une composante essentielle d'une démocratie qui fonctionne ». Des journalistes scientifiques ont été recrutés, d'autres ont été redécouverts par leur rédaction, les pages sciences ont été plus lues que jamais, de nouveaux médias scientifiques comme Epsiloon se sont lancés... Des médias généralistes et des écoles de journalisme repensent leur formation aux sciences. « *Nous sommes à un moment particulier, entre crise sanitaire, écologique, énergétique, où peut-être cette révolution culturelle des médias peut s'engager* », espère Yves Sciama.

Défiance

Le thème Science et crise s'est imposé de lui-même, rappelle Audrey Mikaëlian de l'AJSPI. Mais ce mot *crise* est-il le plus juste pour décrire ce que nous vivons ? La crise environnementale a émergé il y a une cinquantaine d'années et durera des millénaires. Et la crise sanitaire semble devenir une nouvelle normalité : les coronavirus ne vont pas disparaître. Reste que la période actuelle a les caractéristiques d'une crise : politisation et conflictualité des débats, défiance envers les politiques et les médias. Quelle sont les implications de ces crises pour les scientifiques et les journalistes ? Comment favoriser un débat public éclairé et factuel ? Trouver des réponses est le but de cette journée.

Instance scientifique transparente

Bernard Jomier, sénateur et rapporteur de la mission d'information Covid 19, rappelle que le président de la République a parlé de guerre contre le covid. Or, en cas de guerre, le mensonge devient légitime ! Mais une épidémie n'est pas une guerre, la santé publique est même l'antithèse de la guerre. « *Notre commission d'enquête a démontré des mensonges assumés, notamment sur les masques* », rappelle-t-il. Quand le président a voulu avoir l'avis des scientifiques, il s'est retrouvé avec un panorama d'institutions bien trop éclaté ; d'où la création d'un Conseil scientifique. Dans notre pays, le lien entre politiques et scientifiques est mal organisé, trop dépendant de personnalités. Bernard Jomier conseille la création d'une instance scientifique pérenne, transparente, et pas seulement dans le domaine sanitaire.

Raoult analysé

Antoine Bayet est directeur éditorial de la *Revue des médias*[2] qui analyse les mutations des médias. Il revient sur les débuts du confinement, et sur la publication concomitante d'un article scientifique de Didier Raoult sur l'hydroxychloroquine pour traiter le covid. La médiatisation a été immédiate, le nom « hydroxychloriquine » étant prononcé jusqu'à 35 fois par heure. Parallèlement est né le groupe Facebook dédié au culte de Didier Raoult. Il regroupe 400 000 membres en quelques semaines. Entre le 23 au 29 mars, ce groupe a généré 1,8 millions de partages, contre 1,7 millions pour six gros médias réunis (BFM, LCI CNEWS, France Info,

2. https://larevuedesmedias.ina.fr/

Le Monde, Le Figaro) ! Le succès médiatique de Didier Raoult découle surtout d'une stratégie de communication incroyablement efficace.

Comment les médias ont vécu la pandémie

La première table ronde, Les médias face à la crise, rassemblait quatre journalistes de médias très différents : Nicolas Martin, producteur de l'émission La méthode scientifique sur France Culture, Mathilde Fontez, rédactrice en chef d'Epsiloon, Céline Pigale, directrice de la rédaction de BFMTV, et Hervé Morin, responsable du cahier Sciences du Monde.

L'animateur, Yves Sciama, leur demande à chacun ce qui a été le plus difficile à gérer pendant la crise, et une leçon qu'ils ont apprise. Pour Mathilde Fontez, une crise comme le covid est un cauchemar pour un mensuel : les délais de publication sont de trois semaines, en plus du temps d'enquête et d'écriture. « *Tout va très vite, comment ne pas faire un article périmé ? On a joué la carte du décryptage, des articles de synthèse. Cela reste très frustrant de ne pas raconter ce qui se passe* ».

A l'inverse, Nicolas Martin a été « le nez dans le guidon » avec une chronique quotidienne de suivi de l'épidémie. Comment y voir clair ? « *On voyait tomber les études les unes après les autres sans voir où ça allait. On a travaillé jour après jour, étude après étude, en montrant l'évolution de la recherche. On a montré la différence entre la science et la recherche, c'est-à-dire la science en train de se faire. Parfois elle se contredit, ou évolue, et ça a semé de la confusion, entretenue par quelques personnalités et quelques médias* ». Pour la première fois de sa carrière, il a fait relire ses chroniques par des spécialistes, pour ne pas laisser la moindre erreur. Il a choisi de répondre aux mails des auditeurs (hors insultes), et subi un burn out.

Pendant l'épidémie, l'équipe de huit personnes du cahier Science et médecine du journal Le Monde a fait ce qu'elle fait habituellement chaque semaine... puissance 10. La dernière semaine d'avril 2020, 500 articles scientifiques étaient mis en ligne, auxquels il fallait ajouter les avis des autorités de santé, des agences sanitaires, du conseil scientifique, la littérature réglementaire... Hervé Morin a appris qu'on peut faire du journalisme de qualité à partir des *preprints*... et qu'il faut parfois se méfier des articles validés par les grandes revues. « *L'incertitude est inconfortable mais elle est parfois instrumentalisée par les politiques pour masquer une impréparation, il faut pouvoir le dire* », souligne-t-il. « *Il faut des journalistes scientifiques dans les rédactions* ».

Pour Céline Pigalle, cette crise a été stupéfiante par son intensité et sa durée : on n'a parlé que de ça pendant des mois. BFMTV ne disposait que d'une seule journaliste spécialiste de santé, elle a désormais un service de 10 personnes. Céline Pigalle a été frappée par l'envie du public de croire à certaines *fake news*. « *Essayer de convaincre par les faits ne marche pas pour une partie de la population* » observe-t-elle. Elle justifie le traitement par BFMTV du cas Didier Raoult par le fait qu'il était, au début, considéré comme sérieux. Puis, il est devenu un personnage de cette

crise qu'on ne peut pas ignorer, d'où une enquête sur lui. Il a une dimension sociologique. « *Pourquoi ne pas l'avoir confronté à un journaliste scientifique, ou à une expertise ?* », rétorque Nicolas Martin. « *En tant que journalistes, on ne veut pas faire de l'eau tiède, on veut trouver des gens intéressants, possiblement minoritaires, mais qui ont des choses à apporter* », constate Yves Sciama. « *Comment gérer cette prise de risque ?* ». Pour Mathilde Fontez, « *le journalisme scientifique est un journalisme de l'information disponible, contrairement à d'autres domaines. Mais comment s'y retrouve-t-on, comment distingue-t-on une mauvaise publication ? C'est la difficulté de notre métier et c'est notre expertise* ». Nicolas Martin dénonce : « *les chercheurs comme Raoult et Montagnier font un mal colossal à l'information scientifique. Quand on montre que leurs propos ne sont pas scientifiquement solides, on se prend des volées d'insultes* ».

Comment lire un article scientifique ?

Gary Dagorn et Mathilde Damger, journalistes aux Décodeurs du journal Le Monde, ont publié un article expliquant comment lire une publication scientifique. Il s'adresse au grand public de bonne volonté, qui essaie de lire directement à la source des articles qui ne sont pas écrits pour lui. Dès mi-2020, des gens s'échangent des publications scientifiques, qui deviennent des armes rhétoriques. Des articles sans valeur scientifiques sont ainsi largement citées. L'idée du journal Le Monde : lutter contre la starification (le scientifique chevalier blanc), la citation hors contexte, le *cherry-picking* (sélectionner uniquement les travaux qui confirment ce que l'on pense). Cet article pédagogique explique, à partir d'un exemple, ce qu'on trouve dans une publication scientifique, les types de publications, la démarche scientifique, quelques éléments vérifiables d'une publication. Il comporte aussi quelques conseils de lecture, et les spécificités selon les disciplines.

Fact-checking

Florian Gouthière, journaliste scientifique à Libération, nous parle du *fact-checking* (vérification de ce qui est dit dans les médias) en temps de crise. Dès le 22 mars, jour de l'annonce du confinement, Florian Gouthière alertait sur un article de Didier Raoult. Il rappelle que le travail scientifique est collectif, cumulatif (chaque découverte s'ajoute aux autres) et correctif. Le *fact-checker* en science fait état des connaissances et des consensus à une date donnée. Le *fact-checking* comporte des écueils : certains sont inutiles (sur des sujets auxquels personne ne croit), d'autres à côté de la plaque, d'autres contreproductifs, mettant en avant des fausses informations peu relayées. « *Problème : de nouvelles données peuvent rendre obsolète une analyse pourtant pertinente à un moment donné* », souligne Florian Gouthière. « *Il faut faire comprendre qu'on n'avait pas tort !* » En temps de crise, le public cherche des certitudes. Or, le *fact-checking* ne sert pas à ça. Il est surtout utile pour armer d'arguments les gens qui souhaitent argumenter avec leurs proches. En se

souvenant toujours que des affirmations extraordinaires nécessitent des preuves extraordinaires.

Dans la tête des complotistes

William Audureau, journaliste aux Décodeurs du journal Le Monde, nous emmène « dans la tête des complotistes » selon le titre de son ouvrage, fruit de son étonnement en découvrant l'émergence de fausses informations. Comment comprendre ce phénomène de diffusion massive de fausses informations ? Pourquoi les articles de vérification restent souvent vains ? « *Le complotisme est un système d'autodéfense intellectuelle incroyablement imperméable à la contradiction* », constate-t-il. Pourtant, le discours se veut très scientifique, il mime l'expertise. Beaucoup de gens qui rentrent dans ces sphères tombent dans l'ultracrépidarianisme : la croyance de pouvoir facilement se prononcer sur des sujets complexes qu'on ne maîtrise pas.

On devient complotiste car on ne comprend pas. L'épidémie a montré la science non comme un stock de connaissances, mais comme une production qui tâtonne, un processus. Or, il existe une intolérance à l'incertitude chez certaines personnes. « *Les complotistes ne quittent pas la science institutionnelle par rejet de se promesses, mais au contraire pour aller vers une science considérée comme plus pure, moins complexe. Ils ont le sentiment de réparer la science* », conclut Wlliam Audureau.

Comment parler aux complotistes ? Il ne faut pas être insultant, agressif ou méprisant. Préparer un terrain fertile pour la discussion, apaisé, essayer de trouver un terrain d'entente, et regarder ensemble certains points, par exemple lire un article scientifique, pour co-construire la connaissance. « *Ça demande beaucoup de patience* », reconnaît-il. « *Il faut surtout rétablir la confiance, car le plus grand carburant du complotisme est la défiance* ».

Au cœur du maelström, les scientifiques dans les médias

S'ensuit une table ronde avec l'épidémiologiste et statisticienne Dominique Costagliola, les épidémiologistes Arnaud Fontanet et Mircea Sofonea, et le mathématicien Stéphane Dhersin. Yves Sciama leur demande à chacun ce qui les a surpris dans le fonctionnement des médias, et un conseil pour des collègues plus novices médiatiquement.

Dominique Costagliola s'étonne qu'on lui ait souvent demandé de répéter la même chose dans différents médias. Elle refuse de donner son point de vue à chaque nouveauté, alors même qu'elle n'a pas eu le temps de se faire un avis. Il faut rester vigilant sur le type d'émission, et savoir qui seront les autres invités. Mircea Sofonea regrette que les médias exploitent parfois une différence de forme, pour la transformer en désaccord de fond. Cela donne une mauvaise image des chercheurs. Il conseille aux collègues (notamment débutants) de répondre en binôme aux médias, pour se compléter, reformuler, trouver le bon exemple...

Pour Stéphane Dhersin, les personnes compétentes sur cette épidémie sont celles qui comprennent les mathématiques (statistiques, modélisations). Or, les médecins ont plus de légitimité sur les plateaux de télévision. Il déplore l'attente démesurée des journalistes, par exemple savoir quand sera le pic. « *Au début, on se fait piéger, ils interprètent ce qu'on dit quand on n'est pas en direct* ». Malgré cela, « *il est important que ceux qui comprennent les articles scientifiques aillent dans les médias pour les expliquer, donner les limites, afin que l'information diffusée soit la plus juste possible* », encourage Stéphane Dhersin.

Arnaud Fontanet, quant à lui, a été surpris par le caractère immédiat et éphémère de la nouvelle. Quand il ne peut pas répondre le jour même, mais le lendemain, ça n'intéresse déjà plus les médias. Il conseille aux scientifiques de rester sur leur domaine, et d'être complet dans ses explications même si le journaliste s'impatiente.

Qui est légitime à s'exprimer ? Stéphane Dhersin répond s'il a pu lire les publications. Dominique considère qu'il faut intervenir seulement sur les sujets sur lesquels on se sent compétent. Elle-même a été critiquée par certains car elle n'est pas médecin, malgré le fait qu'elle fasse de la recherche clinique depuis 40 ans. Elle s'est pris des tombereaux de haine. Mircea Sofonea n'hésite pas à envoyer vers des collègues d'autres domaines. Il appelle les médias à dresser une topologie des disciplines relatives aux maladies infectieuses : virologue, épidémiologiste, infectiologue, médecin de santé publique... « *Faut-il créer une déontologie de la communication médiatique ?* » demande Mircea Sofonea. Un scientifique ne sera jamais sanctionné pour s'exprimer en dehors de son champ de compétence. Pour Arnaud Fontanet, être membre du Conseil scientifique a compliqué sa relation avec les médias. Stéphane Dhersin rappelle que les scientifiques peuvent estimer l'impact de mesures, mais c'est aux politiques de prendre les décisions.

Que peuvent faire les institutions ? Du *médiatraining*, si possible avant une crise. Pour Dominique Costagliola, cette crise a révélé un manque de cadre déontologique par rapport à la communication de la part des scientifiques. Un directeur de recherche a une crédibilité et peut dire n'importe quoi. Mais il y a aussi la liberté académique. Où se trouve l'équilibre raisonnable ?

Yves Sciama regrette que les institutions ne réagissent pas face à la désinformation par leurs salariés. Jean-Stéphane Dhersin pense que les journalistes les invitent en connaissance de cause, pour le buzz. Arnaud Fontanet remarque que les journalistes et les scientifiques ont beaucoup de choses en commun, notamment la vérification des faits. « *J'ai vu des journalistes faire un travail remarquable et être source d'informations très utiles pour nous* », souligne-t-il. « *Malheureusement, 5 % de personnes avec des pratiques non conformes, côté scientifiques comme côté journalistes, ont entaché la couverture de cette crise* ».

Crises au ralenti... comment être entendus

« *Si la crise du Covid a suscité des débats électriques et mis la société au bord de la crise de nerf, elle a au moins été gérée* », rappelle Yves Sciama lors de nouvelle table ronde. « *Les politiques s'en sont emparés, la société en a débattu, elle n'a pas été niée ou ignorée. La plupart des crises environnementales n'ont pas eu cette* chance *: crise de la biodiversité, crise climatique... ne sont pas gérées. Qu'est-ce que les scientifiques peuvent faire de plus que de produire du savoir ?* »

Pour Bruno David, paléontologue et président du Muséum national d'histoire naturelle, « *la pandémie est une occasion magistralement manquée de faire comprendre ce qu'est la science, comment elle avance, avec ses hypothèses, ses doutes, parfois des remises en question* », estime-t-il. La science n'est ni une opinion, ni une croyance, ni une idéologie. Il ne faut pas non plus confondre la science (ce que l'on connaît, qui est acquis) et la recherche (ce qu'on ne connaît pas encore vraiment). Or, pendant la pandémie, les gens attendaient quelque chose de stable comme la science. « *Le rôle du Muséum est de donner aux citoyens les moyens de se forger leurs propres opinions sur des bases rationnelles* », conclut Bruno David.

« *J'ai l'impression depuis deux ans que nos collègues épidémiologistes et spécialistes de santé ont vécu en accéléré ce qu'on ressent depuis 30 ans en sciences du climat* », observe Valérie Masson-Delmotte, climatologue et co-présidente du groupe 1 du GIEC ; notamment la parole donnée à des personnalités à fort égo, une incompréhension des enjeux de la part de journalistes économiques ou politiques, le faible poids des journalistes scientifiques dans les rédactions, le manque de formation sur la transmission aéroportée du covid comme sur le climat, le déni de gravité de la situation, l'aversion au changement, le poids disproportionné accordé aux voix discordantes (surtout quand c'est rassurant) sont similaires. Valérie Masson-Delmotte regrette aussi une difficulté institutionnelle de la communauté scientifique à réagir. Les meilleures initiatives viennent souvent de jeunes chercheurs, par exemple *climate feedback* permettant de faire rapidement du *fact-checking* de journaux. Dernier enjeu de taille : comment se former collectivement en réfléchissant à la manière dont on s'exprime, dont c'est reçu, comment on peut travailler avec les journalistes, avoir un retour d'expérience... « *Tout cela se travaille, s'apprend avec les spécialistes de la communication scientifique. Il faut que les scientifiques se sentent légitimes à s'exprimer ans les médias* ». La désinformation ne vient pas seulement de gens de bonne foi, mais aussi de *marchands de doute*, qui la diffusent insidieusement dans la société pour servir certains intérêts.

Aurore Gély-Pernot, chercheuse à l'École des hautes études de santé publique travaille sur les perturbateurs endocriniens (PE), susceptibles de poser des problèmes de santé à long terme. « *C'est une épidémie silencieuse, due à des molécules qui titille notre système endocrinien* », explique-t-elle. Les PE sont partout, les interdire est

impossible, alors qu'on y est exposés tous les jours. Certains comme le chlordécone persistent des décennies dans la nature. Il faudrait une régulation européenne.

Wolfgang Kramer, écologue, modélisateur et membre du GIEC est en colère contre les médias. « *Nous sommes des lanceurs d'alerte robustes, réfléchis, calmes depuis des décennies, mais nous n'avons pas réussi à communiquer les points essentiels* », regrette-t-il. « *Quels moyens avons-nous pour faire passer le message ?* » D'où une proposition provocatrice de faire la grève, arrêter de faire de la science. « *À un moment, il faut s'interroger : ce que nous faisons suffit-il ? Nous devons réfléchir avec les journalistes sur notre manière de communiquer* ». Valérie Masson-Delmotte suggère de créer des ponts entre jeunes scientifiques et jeunes journalistes.

Pour Bruno David, « *si on entre dans l'action, on perd une partie de notre crédibilité* ». Le Muséum s'adresse aux politiques via les médias, et au public par l'émerveillement, en l'amenant à regarder à la beauté du monde (par exemple lors de l'exposition sur l'océan), donc à être plus sensibles à l'environnement. C'est la sensibilisation par l'émotion. Et l'émotion fixe la mémoire.

« *Comment communiquer sans angoisser* », demande une personne de la salle. Wolfgang Kramer s'interroge sur l'injonction à ne pas être anxiogène, à dire que les petits pas vont dans le bon sens. « *Si on est sur le Titanic, on ne dit pas qu'un petit virage va suffire ! On ne peut pas adoucir la gravité de la situation sous le prétexte de ne pas être anxiogène* ». Valérie Masson-Delmotte lit les travaux des spécialistes de psychologie expérimentale, de science du comportement, qui éclairent la manière dont des connaissances factuelles sont perçues. Cela crée des émotions qui ressemblent à celles du deuil, pouvant créer de la paralysie et du déni, ou au contraire de la lucidité et de l'action. Elle conseille de montrer où on peut agir. Les jeunes sont inquiets non parce que le changement climatique est anxiogène, mais qu'ils ont l'impression que les adultes n'agissent pas. Agir est moins anxiogène que ne pas parler !

Les Français et la science

Michel Dubois, sociologue au CNRS, a enquêté sur les français et la science en 2021 [3], avec Martin Bauer et Pauline Hervois. Cette enquête clarifie des confusions fréquentes dans les débats, par exemple entre science et expertise, entre science et applications, et entre le grand public et les minorités mobilisés. « *Il n'y a pas de défiance majoritaire à l'égard des scientifiques, il y a au contraire une confiance de principe* », observe Michel Dubois. « *Mais une majorité considère désormais que la science apporte autant de bien que de mal* ». Sur le coronavirus, le public fait confiance aux médecins et universitaires, mais pas aux journalistes. Le grand public ne distingue pas les journalistes scientifiques des autres. L'indépendance des journalistes, notamment, est remise en doute. La manière dont les gens votent (donc leur idéologie) joue énormément sur cette confiance. La relation des français à la science

3. http://www.science-and-you.com/fr/sondage2021

est singulière par rapport à d'autres pays européens, ou à l'Amérique du Nord. Nous avons un rapport *utilitariste* au savoir scientifique, et un doute sur l'indépendance des chercheurs par rapport à des intérêts industriels.

Relations entre science et politique

Cédric Villani, mathématicien, député et président de l'OPECST [4], nous raconte l'histoire des relations entre science, crise et politique. La seconde guerre mondiale donne naissance une science triomphante, avec la mise au point d'une arme destructrice qui change totalement la donne, puis la création de grands programmes scientifiques. À côté, la crise du covid paraît extraordinairement confuse, et la confiance en la science semble ébranlée. Pourtant, quand on se replonge dans le passé, on observe aussi un désordre considérable et des tensions permanentes entre scientifiques et militaires. D'où ce constat : la science sous pression fait des erreurs et engendre des incompréhensions.

Mais notre rapport à l'information a changé. Les débats normaux dans un contexte scientifique sont exposés au public, avec parfois des effets ravageurs. Des scientifiques sont devenus *superstars* en s'adressant directement aux citoyens via les réseaux sociaux. Avec des dérives : le narratif du scientifique rebelle contre son institution est plus efficace que les arguments scientifiques.

« *Loin d'effacer les différents politiques, cette crise les a renforcés* », constate aussi Cédric Villani. Pas de trêve entre politiques pour faire face au covid. Le pays le plus admiré dans sa gestion de crise, la Corée du Sud, est institutionnellement préparé pour faire face. En cas de pandémie, un organisme de santé public prend des pouvoirs temporaires et se met aux manettes. En France, au contraire, on avait une dizaine d'organisations, créant une cacophonie. La confiance des citoyens était plus orientée par leur confiance dans le gouvernement que dans la science. « *J'ai pu voir combien il était pratique d'avoir un organisme pré-existant, rassemblant scientifiques et politiques, opérationnel et indépendant* », souligne le mathématicien, qui appelle à organiser les crises en amont.

Caroline Lachowsky, productrice à Radio France internationale, conclut la journée en nous rappelant les points clé abordés. Un grand merci à Guillaume Monnain de l'agence de communication Akenium pour ses dessins en live lors de cette journée.

4. Office parlementaire d'évaluation des choix scientifiques et technologiques.

Les réseaux d'automates booléens au cœur du calcul naturel

Kévin Perrot [1] et Sylvain Sené [2]

Depuis la naissance de la science informatique moderne dans les années 1930-1940, la biologie n'a eu de cesse d'inspirer ses développements sur les plans théorique et appliqué (réseaux de neurones, automates cellulaires, calcul moléculaire...). Réciproquement, depuis les années 1960, l'informatique joue un rôle fondamental dans les avancées en sciences du vivant. Il n'y a qu'à voir les innovations réalisées en algorithmique des séquences, en biologie structurale et en phylogénie pour se rendre compte de l'influence de notre science sur la biologie. Au delà de ces champs de recherche, Jacob et Monod [9] ont mis en avant le caractère essentiel des régulations biologiques, et donc de la transmission d'information. Leurs travaux ont ouvert la voie à un nouveau domaine, ancré en mathématiques et en informatique et visant la compréhension du vivant, s'attachant à analyser la manière dont les entités biologiques interagissent pour engendrer les phénomènes observés. Du point de vue de l'informatique, il s'agit de comprendre comment ces transmissions d'information engendrent les calculs opérés par les organismes étudiés, ce qui correspond exactement à la définition originelle de *bio-informatique* telle qu'introduite par Hesper et Hogeweg [8] : « *l'étude des processus informatiques dans les systèmes biologiques* ».

Dans ce cadre, nous proposons un survol introductif du modèle des réseaux d'automates booléens. Après avoir défini l'objet mathématique, nous parcourrons son histoire et les liens qu'il entretient avec l'informatique et la biologie. Une première série de résultats théoriques montre que les rétroactions jouent un rôle moteur dans

1. Université d'Aix-Marseille, CNRS, LIS, Marseille.
2. Université publique, Marseille.

DOI:10.48556/SIF.1024.20.171 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 171–182

la complexité des phénomènes de régulation, et une seconde série témoigne de la difficulté à analyser la dynamique de tels systèmes.

Les réseaux d'automates booléens

Le modèle mathématique des réseaux d'automates capture de façon générale les systèmes composés d'entités en interaction (qui incluent donc les phénomènes de régulation). Un réseau d'automates peut être vu comme un ensemble de machines exécutant des calculs (les automates), qui dépendent des informations qu'elles échangent avec d'autres machines du réseau. Dans le contexte de cet article, à des fins de simplification, nous nous restreignons aux réseaux d'automates booléens, dans lesquels les informations transmises et retenues par les automates, sont binaires (0 ou 1). Plus formellement, dans un *réseau d'automates booléens* de taille n (composé de n automates indexés par $[n] = \{0, \dots, n-1\}$), chaque *automate* $i \in [n]$ possède un *état booléen* $x_i \in \{0,1\}$, et met à jour son état selon une *fonction locale* $f_i : \{0,1\}^n \to \{0,1\}$. L'affectation d'un état à tous les automates du réseau donne une *configuration* $x \in \{0,1\}^n$. Les fonctions locales définissent les influences des automates les uns sur les autres, capturées par un *graphe d'interaction*. Il s'agit d'un graphe orienté composé d'un sommet par automate, avec un arc de l'automate i vers l'automate j lorsque l'actualisation de l'état de j, c'est-à-dire $f_j(x)$, dépend effectivement de l'état de i, c'est-à-dire x_i. On ajoute généralement des signes à ces dépendances, pour spécifier la nature de l'influence : un signe $+$ sur l'arc (i, j) spécifie que l'automate j tend à mimer l'état de l'automate i ; un signe $-$ souligne quant à lui que j tend à prendre l'état opposé de celui de i. Par exemple, pour la fonction $f_j(x) = x_i \vee \neg x_j$, le graphe d'interaction correspondant a un arc positif (i, j) et une boucle négative (j, j). Dans le cas d'une fonction $f_j(x) = x_i \oplus x_j$ (avec $\oplus$ le *ou exclusif*), le graphe d'interaction contient l'arc (i, j) et la boucle (j, j) signés à la fois positivement et négativement.

La donnée des fonctions locales est une description statique (syntaxique) du réseau. Pour obtenir un système dynamique discret sur l'espace $\{0,1\}^n$ qui permettrait d'appréhender la transmission d'information et les calculs opérés, il faut en outre lui associer un *mode de mise à jour*. Ce dernier détermine la manière dont les automates mettent à jour leur état dans le temps discret. Parmi les modes de mise à jour les plus présents dans la littérature, se trouvent :

— le mode *parallèle*, qui est déterministe et qui actualise l'état de chaque automate à chaque étape de temps ;

— le mode parfaitement *asynchrone*, qui est non-déterministe et qui actualise l'état d'un automate à chaque étape de temps.

Le couple (f, μ), avec f un réseau d'automates et μ un mode de mise à jour, définit un système dynamique $f_\mu : \{0,1\}^n \to \{0,1\}^n$, qui peut être représenté par un *graphe de transition*. Cette description dynamique (sémantique) est un graphe orienté qui a

pour ensemble de sommets $\{0,1\}^n$ (l'espace des configurations), et un arc de x vers y, appelé transition, lorsque $f_\mu(x) = y$ dans le cas déterministe, ou $y \in f_\mu(x)$ dans le cas non-déterministe. On a par exemple :

— $f_{\mathsf{p}}(x) = (f_0(x), \ldots, f_{n-1}(x))$ pour le mode parallèle p ;
— $f_{\mathsf{a}}(x) = \{(x_0, \ldots, x_{i-1}, f_i(x), x_{i+1}, \ldots, x_{n-1}) \mid i \in [n]\}$ pour le mode parfaitement asynchrone a.

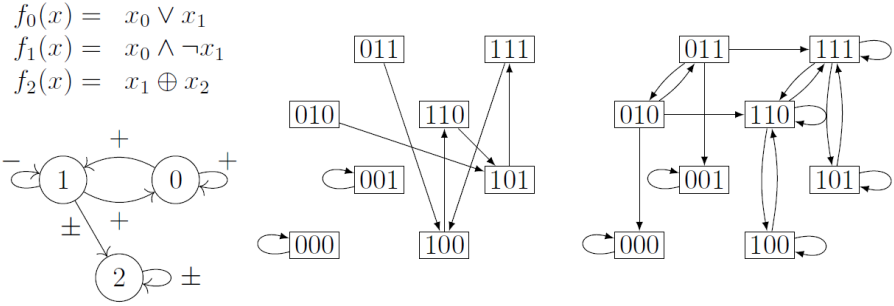


FIGURE 1. Fonctions locales d'un réseau d'automates booléens de taille $n = 3$ (à gauche en haut), son graphe d'interaction (à gauche en bas, avec $\pm$ pour les dépendances à la fois positives et négatives), son graphe de transition pour le mode de mise à jour parallèle (au centre), et son graphe de transition pour le mode de mise à jour parfaitement asynchrone (à droite). Dans le mode parallèle la configuration 011 a pour image $f_{\mathsf{p}}(011) = 100$ où tous les automates changent d'état (en effet, $f_0(011) = 0 \vee 1 = 1$, $f_1(011) = 0 \wedge \neg 1 = 0$ et $f_2(011) = 1 \oplus 1 = 0$), ce qui se traduit dans le mode parfaitement asynchrone par trois transitions possibles $f_{\mathsf{a}}(011) = \{111, 001, 010\}$ où chacun des automates peut mettre à jour son état de façon non-déterministe.

Un exemple est proposé en figure 1. Sur cette figure, on observe que les dynamiques parallèle et asynchrone sont bien différentes, ce qui souligne la sensibilité des réseaux aux modes de mise à jour. Dans la dynamique parallèle, chaque configuration a une et une seule transition possible [3] alors que, dans le cas asynchrone, chaque configuration peut avoir plusieurs transitions possibles, comme par exemple la configuration 010. Ce qui nous intéresse en termes de transmission d'information et de calcul, ce sont les chemins dans la dynamique (trajectoires), et ce vers

3. Cette propriété est vérifiée par tout mode de mise à jour déterministe.

quoi ces chemins mènent (comportements limites, qui sont les composantes fortement connexes terminales [4] du graphe de transition). Deux types de comportement limite nous intéressent particulièrement : les configurations stables et les oscillations stables, communément appelés points fixes et cycles limites dans le cas déterministe. Dans l'exemple de la figure 1, on observe que la dynamique selon le mode parallèle possède deux points fixes $f_{\mathsf{p}}(000) = 000$ et $f_{\mathsf{p}}(001) = 001$, et un cycle limite de période quatre $f_{\mathsf{p}}^4(100) = f_{\mathsf{p}}^3(110) = f_{\mathsf{p}}^2(101) = f_{\mathsf{p}}(111) = 100$. On remarque également que les points fixes du mode parallèle sont aussi des points fixes du mode asynchrone, car aucun automate ne change d'état lors de l'application de sa fonction locale. Dans le mode asynchrone, on a de plus une oscillation stable composée des configurations $100, 110, 101, 111$ (il s'agit ici du même ensemble de configurations que le cycle limite du mode parallèle, mais ce n'est pas toujours le cas). Une trajectoire est une séquence de configurations obtenues par des transitions successives dans le système dynamique, par exemple $(010, 011, 111, 111, 101, 111, 110)$ pour le mode asynchrone sur la figure 1.

Un modèle de calcul naturel

Du point de vue de l'informatique et des mathématiques, les réseaux d'automates sont une famille de systèmes dynamiques discrets, qui peut être utilisée pour capturer des phénomènes biologiques. Ces phénomènes biologiques sont en général décrits à la suite d'expérimentations en laboratoire « humide ». Cette première description peut être assez éloignée du formalisme des réseaux d'automates booléens (même si, en pratique, le langage utilisé est proche de la logique propositionnelle), ce qui nécessite une étape de *traduction*. Par exemple, dans le cadre de la modélisation des réseaux de régulation génétique, chaque gène est représenté par un automate (pour faire simple) dont l'état booléen représente son état d'expression (1 pour activé, 0 pour inhibé). Cet état représente le fait qu'il soit transcrit (ou non) en ARN messager puis traduit (ou non) en protéines. L'expression d'un gène dépend essentiellement de l'expression des autres gènes de son réseau (plus précisément des taux de concentration des protéines produites). On peut alors décrire des observations sous la forme « si les gènes *A* et *B* sont exprimés, alors le gène *C* est inhibé ». Lorsque de telles spécifications sont incomplètes (ce qui est toujours le cas), on peut construire plusieurs réseaux d'automates qui les satisfont. Des informations supplémentaires (observations temporelles de rapidité des réactions en chaîne, différents paliers d'activation...) peuvent permettre de réduire cet ensemble de modèles. On a alors deux types de *modèle* : d'un côté le modèle mathématique des réseaux d'automates, de l'autre des « modèles biologiques » correspondant à des instances du modèle mathématique. L'intérêt de l'analyse de la dynamique limite de tels modèles réside dans le fait que les configurations et les oscillations stables du réseau d'automates sont

4. C'est-à-dire desquelles ne sort aucune transition.

assimilables aux phénotypes (types cellulaires, tissus...) et aux rythmes biologiques (circadien, cardio-respiratoire, cycle cellulaire...) émergeant du système.

Ce qui est certain, c'est qu'aucun modèle biologique ne pourra représenter parfaitement la réalité. En effet, étant une abstraction d'observations incomplètes de la réalité, on ne peut espérer que ces modèles capturent la totalité des paramètres en jeu dans le système réel. Comme l'écrivent Delahaye et Rechenmann [5] : « *Tout modèle est faux, et c'est très bien !* ». La qualité d'un modèle s'évalue à son utilité plutôt qu'à sa correction au regard de la réalité. Les modèles des réseaux d'automates en sont de parfaits exemples. La simplicité de leur définition force leur incorrection vis-à-vis des systèmes qu'ils modélisent, de même qu'elle est grandement utile car elle fait porter l'attention sur l'essence des interactions et des transmissions d'information qui engendrent la complexité observée. Il s'agit ainsi d'une modélisation qualitative permettant d'énoncer des propriétés générales des systèmes, car cette simplicité « libère » les modèles de nombreux paramètres.

Jusque là, nous avons présenté le modèle des réseaux d'automates au travers d'allers-retours entre la biologie et l'informatique mais, une fois passé au modèle mathématique, il convient de l'étudier en profondeur *per se*. Ces inspirations tirées de phénomènes réels offrent en effet de nouveaux objets aux mathématiciens et informaticiens, posant de nouvelles questions sur la théorie des systèmes dynamiques, la combinatoire, et la science du calcul (en particulier au travers de la complexité et de la calculabilité) : un réseau d'automates associé à un mode de mise à jour peut être vu comme un programme, dont la dynamique est l'exécution d'un calcul. S'inspirer de la phénoménologie, en tirer des modèles de calcul, les étudier en profondeur, puis les utiliser pour mieux comprendre les phénomènes naturels, est ce qui fonde le domaine du *calcul naturel*. D'un côté, il s'agit d'envisager la nature comme un ensemble de processus calculatoires à différentes échelles, et d'utiliser la science informatique pour les analyser ; de l'autre, il s'agit d'observer la nature pour découvrir et imaginer de nouvelles façons de calculer. Cette vision du calcul au sein de la nature, et des aspects naturels du calcul, soulève un intérêt croissant de la communauté scientifique, et nous croyons qu'il s'agit d'une direction prometteuse à la fois pour mieux comprendre le vivant et pour le développement de nouvelles technologies.

Une histoire croisée

Bien que le calcul naturel paraisse être un domaine assez récent, il existait déjà aux fondements de l'informatique moderne. Outre le modèle des machines de Turing inspiré du mathématicien devant sa feuille [24], les réseaux de neurones ont été introduits par McCulloch et Pitts [12] pour modéliser les processus cérébraux, et les automates cellulaires ont été développés par von Neumann et Ulam [14] pour conceptualiser l'auto-réplication observée en biologie. Par ailleurs, ces trois modèles

ont donné naissance à des pans entiers de l'informatique moderne et à leur développement. Les machines de Turing ont directement permis de définir le concept de calculabilité qui est le premier pilier de la science informatique. Les réseaux de neurones de McCulloch et Pitts ont quant à eux notamment donné naissance à la théorie des automates finis [11], et ont participé à celle de l'intelligence artificielle [21, 13]. En souhaitant « *comparer les automates naturels et artificiels* » et « *abstraire la structure logique de la vie* », von Neumann a créé un des premiers modèles mathématiques d'architecture d'ordinateur parallèle. Ces deux derniers modèles sont en fait des cas particuliers de réseaux d'automates (le premier avec des fonctions locales à seuil, le second sur une grille avec une homogénéité spatiale).

À l'instar de l'informatique moderne qui s'est largement inspirée de la biologie, cette dernière fait largement appel à la première depuis que Jacob et Monod ont défendu le choix d'une approche systémique en science du vivant. Les réseaux d'automates sont notamment devenus un modèle central pour la modélisation qualitative en biologie moléculaire depuis la fin des années 1960. Ils possèdent en effet les bonnes propriétés pour capturer les caractéristiques structurelles et dynamiques des systèmes biologiques au niveau de la cellule, comme les réseaux de régulation génétique. Par ailleurs, la compréhension des processus de régulation est un problème abstrait que la biologie ne peut résoudre par elle-même. Cette observation fondamentale a été faite en parallèle par Kauffman et Thomas dans le contexte de la régulation génétique : ils soulignent dans [10, 22], l'inadéquation des techniques expérimentales pour aborder le problème dans son entièreté. En effet, bien qu'elles permettent d'acquérir une fine connaissance des éléments biologiques étudiés (sujette à l'interprétation humaine, aux erreurs d'observation, aux biais statistiques...), pour des raisons de complexité notamment, elles empêchent de mettre ces connaissances en perspective et d'obtenir suffisamment de recul pour comprendre les causalités en jeu au sein des combinaisons de régulations. Sur la base de ce constat, Kauffman et Thomas introduisirent des premiers modèles de réseaux de régulation génétique en se fondant sur les réseaux d'automates booléens. La communauté scientifique de biologie des systèmes a vite compris l'intérêt de ces réseaux, qui sont rapidement devenus le modèle mathématique le plus utilisé et le plus influent dans le domaine de la modélisation qualitative des phénomènes de régulation biologique.

Les liens entre structure et dynamique

La principale force du modèle mathématique des réseaux d'automates booléens est la portée générale des résultats obtenus. En effet, par sa capacité à abstraire de nombreux détails des systèmes réels modélisés, le modèle mathématique offre un large potentiel d'applications. Ces applications peuvent concerner non seulement la vérification de la cohérence de résultats expérimentaux et l'aide au choix des prochaines expériences à mener, mais encore la prédiction du comportement d'un

système. Voici un premier exemple de résultat théorique associé à de potentielles applications.

Théorème 1 ([20])**.** *Si le graphe d'interaction d'un réseau de n automates est acyclique, alors sa dynamique converge en au plus n étapes vers une unique configuration stable.*

Pour le mode parallèle, le théorème 1 conclut que f_{p}^n est une fonction constante, c'est-à-dire qu'il existe une configuration $\tilde{x} \in \{0,1\}^n$ telle que, pour toute configuration x, on a $f_{\mathsf{p}}^n(x) = \tilde{x}$. Dans le mode parfaitement asynchrone, le théorème 1 conclut que f_{a} possède une trajectoire de longueur au plus n depuis toute configuration x vers la configuration $\tilde{x}$. Il peut alors être appliqué pour vérifier la cohérence d'un modèle biologique et aider au choix des expériences à mener : si plusieurs configurations stables sont observées mais que le graphe d'interaction est acyclique, c'est que le modèle est incohérent, et qu'il doit exister des interactions créant des cycles dans le graphe d'interaction. Le théorème 1 permet également de prédire le comportement d'un modèle biologique : si le graphe d'interaction est acyclique et que la dynamique est inconnue, on peut prédire qu'il y aura une unique configuration stable. Les résultats théoriques donnent en outre des pistes d'explication mathématique à des phénomènes observés. Par exemple, le théorème 1 montre que les cycles dans le graphe d'interaction, également appelés *boucles de rétroaction* (où une entité agit, parfois indirectement, sur elle-même), sont des moteurs de la complexité dynamique : sans eux, l'évolution est « simple » car, peu importe la configuration initiale, le système convergera toujours vers la même configuration stable.

Parmi les cycles du graphe d'interaction, on distingue les cycles positifs qui comportent un nombre pair d'arcs négatifs, et les cycles négatifs qui comportent un nombre impair d'arcs négatifs. Sur l'exemple de la figure 1, le cycle entre les automates 0 et 1 est positif (zéro arc négatif), alors que le cycle de l'automate 1 vers lui-même est négatif (un arc négatif). L'état d'un automate est « inversé » lorsque qu'il traverse un arc négatif. En revanche, en traversant un nombre pair d'arcs négatifs, cette information revient « à l'identique » sur l'automate lui-même. Un cycle positif est donc intuitivement « bon » pour la stabilité. Ce lien intuitif entre les cycles positifs et les configurations stables d'un réseau a été conjecturé par Thomas, et il a fallu attendre quelques décennies pour en avoir une démonstration formelle.

Théorème 2 ([23, 19, 15])**.** *Si la dynamique d'un réseau d'automates possède plusieurs configurations stables distinctes, alors son graphe d'interaction contient un cycle positif.*

De façon analogue, dans le cadre asynchrone, il existe un lien entre les oscillations stables dans la dynamique et les cycles négatifs dans le graphe d'interaction.

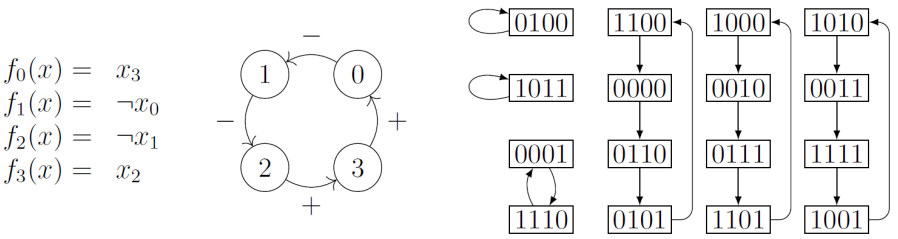


FIGURE 2. Un cycle positif à $n = 4$ automates. Fonctions locales (à gauche), graphe d'interaction (au centre), et dynamique limite (à droite, où l'on peut remarquer que toutes les $2^4 = 16$ configurations sont présentes).

Théorème 3 ([23, 18])**.** *En asynchrone, si la dynamique d'un réseau d'automates possède une oscillation stable, alors son graphe d'interaction contient un cycle négatif.*

Nous avons déjà souligné l'importance de la dynamique limite (les configurations et oscillations stables) pour les applications à la biologie. Les résultats ci-dessus donnent des conditions nécessaires à leur apparition. Pour aller plus loin dans l'étude de la dynamique d'un réseau, il convient de caractériser l'ensemble des comportements limites possibles. C'est ce qu'ont réalisé les auteurs de [7, 6], en commençant par les réseaux d'automates booléens dont le graphe d'interaction est un cycle (positif ou négatif), et pour les modes de mise à jour parallèle et asynchrone. De tels développements constituent une toute première étape dans l'étude d'une version discrète du seizième problème de Hilbert, et font déjà apparaître des formules non-triviales liées aux diviseurs de la taille du cycle (fonction indicatrice d'Euler, formule d'inversion de Möbius). Par exemple, pour un cycle positif de $n = 4$ automates, la dynamique parallèle possède 2 points fixes, 1 cycle limite de période 2, et 3 cycles limites de période 4 (voir figure 2).

L'influence des modes de mise à jour sur la dynamique limite reçoit également une attention particulière, et il s'agit d'un problème complexe. En partant du mode de mise à jour asynchrone et en autorisant une seule transition synchrone (c'est-à-dire qui met à jour l'état de plusieurs automates simultanément), la dynamique limite peut changer drastiquement (modification des trajectoires, et fusion ou destruction d'oscillations stables) [16]. Cette direction de recherche s'attaque à la sensibilité au synchronisme des réseaux d'automates ; elle est d'une grande importance pour les applications à la biologie où l'ordre de mise à jour des entités fait toujours l'objet de débat dans la communauté. On peut néanmoins affirmer que les configurations stables (points fixes dans les modes de mise à jour déterministes) bénéficient d'une

plus grande robustesse au synchronisme, car elles sont invariantes [5] pour le mode parallèle, le mode asynchrone et tous les modes de mise à jour bloc-séquentiels (où les automates sont mis à jour selon une partition ordonnée [6] de $[n]$). Cette robustesse au mode de mise à jour des configurations stables n'est toutefois pas absolue, car des modes de mise à jour plus complexes peuvent engendrer des configurations stables que l'on ne retrouve pas dans les modes classiques que sont le parallèle et l'asynchrone (par exemple avec les modes de mise à jour bloc-parallèles [17]). En résumé, il reste beaucoup de choses à comprendre sur la sensibilité des dynamiques limites à la façon dont sont mis à jour les automates au cours du temps !

L'éclairage de la théorie de la complexité

Dans cette section nous allons considérer le mode de mis à jour parallèle uniquement, et aller plus loin dans le comptage du nombre de points fixes des réseaux d'automates booléens. Des relations ont été établies entre la structure du graphe d'interaction et le nombre de points fixes de la dynamique parallèle. On parle de *bornes structurelles*.

Théorème 4 ([1, 2])**.** *Soit p le nombre de points fixes d'un réseau d'automates booléens f dont le graphe d'interaction est G. Alors $p \leqslant 2^{\tau(G)}$, où $\tau(G)$ est le nombre minimum de sommets à retirer pour rendre le graphe G acyclique. Si de plus il n'y a aucun arc signé à la fois positivement et négativement (on parle de réseau* monotone*), alors $\nu(G)+1 \leqslant p$, où $\nu(G)$ est le nombre maximum de cycles disjoints que l'on peut trouver dans G.*

En utilisant les signes des arcs du graphe d'interaction, ces travaux donnent des bornes plus précises. Par exemple, pour la borne supérieure, l'acyclicité de G peut être remplacée par l'absence de cycle positif (c'est-à-dire $p \leqslant 2^{\tau^+(G)}$, où $\tau^+(G)$ est le nombre minimum de sommets à retirer pour supprimer tous les cycles positifs de G, en laissant possiblement des cycles négatifs).

Ces bornes sont *optimales*, dans le sens où elles sont atteintes par certains réseaux d'automates booléens. On a par exemple le réseau « identité » de taille n, dont les fonctions locales sont $f_i(x) = x_i$ pour tout $i \in [n]$, dont le graphe d'interaction G est composé de n boucles, et qui possède $2^{\tau(G)}$ points fixes : il faut retirer tous les sommets de G pour supprimer tous les cycles, et toutes les configurations sont des points fixes. Pour d'autres réseaux en revanche, ces bornes structurelles sont assez

5. C'est-à-dire que pour un réseau d'automates donné, on aura le même ensemble de configurations stables pour chacun de ces modes de mise à jour.

6. Par exemple dans un réseau avec $n = 6$ automates, on peut mettre à jour les automates 0 et 1, puis les automates 2, 4 et 5, puis enfin l'automate 3, ce qui correspond à la partition ordonnée $(\{0,1\},\{2,4,5\},\{3\})$.

éloignées. Par exemple, en ajoutant un automate supplémentaire au réseau « identité » précédent, d'indice n et de fonction locale $f_n(x) = \neg x_n$, la valeur de $\tau(G)$ ne change pas mais il n'y a plus aucun point fixe car, pour toute configuration, l'état de l'automate n change. On obtient donc un réseau de $n+1$ automates avec $\tau(G) = n$, mais aucun point fixe. Ces bornes structurelles sont donc très générales, mais parfois imprécises. Savoir pour quels graphes orientés la borne supérieure est atteinte est d'ailleurs un problème fameux dans le domaine du *codage de réseau*, qui permet d'optimiser le trafic internet en mélangeant les contenus des paquets tout en étant capable de les ré-assembler lorsqu'ils arrivent à destination !

Un autre point de vue permet de comprendre pourquoi il est difficile d'affiner les bornes structurelles données par le théorème 4 : la complexité algorithmique. Si l'on ne connaît que le graphe d'interaction signé (avec les signes des arcs) à n sommets, plusieurs réseaux d'automates booléens de taille n sont possibles (et même un très grand nombre car pour chaque automate il y a une quantité doublement exponentielle en n de fonctions locales possibles). Comment calculer le nombre maximum de points fixes possibles, noté $p_{\mathsf{max}}(G)$, parmi tous ces réseaux d'automates booléens ? Existe-t-il un algorithme efficace pour calculer $p_{\mathsf{max}}(G)$ à partir de G ? La réponse est non, à moins que $\mathsf{P} = \mathsf{NP}$, c'est-à-dire à moins que de nombreux problèmes que l'on ne sait actuellement pas résoudre efficacement puissent eux aussi être tous résolus efficacement par « essentiellement » le *même* algorithme [4].

Théorème 5 ([3])**.** *Étant donné un graphe d'interaction signé G, on peut décider en temps polynomial si $p_{max}(G) \geqslant 1$, mais décider si $p_{max}(G) \geqslant k$ est NP-complet pour tout $k \geqslant 2$.*

Le théorème 5 indique que l'on peut décider avec un algorithme efficace (bien que l'algorithme en question soit très compliqué, il est « rapide ») s'il est possible d'avoir un réseau d'automates booléens dont le graphe d'interaction est G et qui a au moins un point fixe [7], mais il est par contre très difficile de décider avec un algorithme s'il est possible d'avoir un réseau d'automates booléens avec au moins deux points fixes sur le même graphe d'interaction. On peut en outre poser une question analogue sur le nombre minimum de points fixes possibles, noté $p_{\mathsf{min}}(G)$, qui est un problème encore plus complexe [8] (et pas complémentaire du problème précédent, car il existe des graphes G avec une différence arbitraire entre $p_{\mathsf{min}}(G)$ et $p_{\mathsf{max}}(G)$).

7. Pour un graphe d'interaction signé G, l'existence d'un réseau d'automates booléens possédant au moins un point fixe est démontré équivalent à la présence d'un cycle positif dans G, ce qui peut être testé en temps polynomial par un algorithme qui a mis environ 20 ans à être découvert et est hautement non-trivial...

8. NEXPTIME est une classe de complexité caractérisant des problèmes très très difficiles à résoudre algorithmiquement (c'est-à-dire qui sont résolus par des algorithmiques très peu efficaces : en temps exponentiel).

Théorème 6 ([3]). *Étant donné un graphe d'interaction signé G, décider si* $p_{min}(G) \leqslant k$ *est* NEXPTIME*-complet pour tout* $k \geqslant 1$.

Puisqu'il est même difficile de calculer le nombre maximum et minimum de points fixes parmi les réseaux d'automates booléens sur *un* graphe d'interaction donné, il est inévitable que des bornes précises et générales fassent appel à des propriétés structurelles très complexes du graphe d'interaction. Les constructions utilisées pour démontrer les théorèmes 5 et 6 possèdent un entrelacement délicat entre les cycles positifs et négatifs des graphes d'interactions, qui esquisse à peine la grande subtilité des propriétés en jeu...

Pour conclure

Au cours de ce bref survol de thématiques de recherches fondamentales sur le modèle des réseaux d'automates, nous avons rencontré des problèmes abstraits aux applications vastes, mais qui s'avèrent très compliqués à résoudre. Ce domaine est tiré par des questions de l'ordre de *graals*, comme la caractérisation des points fixes et des cycles limites faisant écho à l'un des 23 problèmes de Hilbert. Il y a pour l'heure peu de pistes de résolution complète de ces questionnements, mais nous avons vu que la complexité algorithmique apporte un regard nouveau qui *explique* en partie pourquoi il ne semble pas y avoir de réponses simples à ces questions complexes. Le modèle mathématique des réseaux d'automates permet toutefois de progresser par étapes, suivant une approche constructive consistant à commencer par des études restreintes, qui sont petit-à-petit composées pour aller en direction de cas plus généraux. Ainsi, même si les grandes interrogations qui portent le sujet sur le temps long sont toujours ouvertes, des progrès significatifs n'ont eu de cesse d'être actés. Ces résultats nourrissent l'enthousiasme d'une large communauté aux multiples facettes !

Références

[1] Julio Aracena. Maximum number of fixed points in regulatory Boolean networks. *Bulletin of Mathematical Biology*, 70 :1398–1409, 2008.

[2] Julio Aracena, Adrien Richard, and Lilian Salinas. Number of fixed points and disjoint cycles in monotone Boolean networks. *SIAM Journal on Discrete Mathematics*, 31 :1702–1725, 2017.

[3] Florian Bridoux, Amélia Durbec, Kévin Perrot, and Adrien Richard. Complexity of fixed point counting problems in Boolean Networks. *Journal of Computer and System Sciences*, 126 :138–164, 2022.

[4] Jean-Paul Delahaye. P = NP, un problème à un million de dollars ? *Interstices*, 2007.

[5] Jean-Paul Delahaye and François Rechenmann. La simulation par ordinateur change-t-elle les sciences ? *Interstices*, 2006.

[6] Jacques Demongeot, Tarek Melliti, Mathilde Noual, Damien Regnault, and Sylvain Sené. *Automata and complexity*, volume 42 of *Springer Series on Emergence, Complexity, Computation*, chapter On Boolean automata isolated cycles and tangential double-cycles dynamics, pages 145–178. Springer, 2022.

[7] Jacques Demongeot, Mathilde Noual, and Sylvain Sené. Combinatorics of Boolean automata circuits dynamics. *Discrete Applied Mathematics*, 160 :398–415, 2012.

[8] Ben Hesper and Paulien Hogeweg. Bioinformatica: een werkconcept. *Het Kameleon*, 1 :28–29, 1970.

[9] François Jacob and Jacques Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3 :318–356, 1961.

[10] Stuart A. Kauffman. *Current topics in developmental biology*, volume 6, chapter Gene regulation networks: A theory for their global structures and behaviors, pages 145–181. Elsevier, 1971.

[11] Stephen C. Kleene. *Automata Studies*, volume 34, chapter Representation of events in nerve nets and finite automata, pages 3–42. Princeton University Press, 1956.

[12] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Journal of Mathematical Biophysics*, 5 :115–133, 1943.

[13] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969.

[14] John von Neumann. *Theory of self-reproducing automata*. University of Illinois Press, 1966.

[15] Mathilde Noual. *Updating Automata Networks*. PhD thesis, École normale supérieure de Lyon, 2012.

[16] Mathilde Noual and Sylvain Sené. Synchronism versus asynchronism in monotonic Boolean automata networks. *Natural Computing*, 17 :393–402, 2018.

[17] Loïc Paulevé and Sylvain Sené. *Systems biology modelling and analysis : formal bioinformatics methods and tools*, chapter Boolean networks and their dynamics : the impact of updates. Wiley. À paraître.

[18] Adrien Richard. Negative circuits and sustained oscillations in asynchronous automata networks. *Advances in Applied Mathematics*, 44 :378–392, 2010.

[19] Adrien Richard and Jean-Paul Comet. Necessary conditions for multistationarity in discrete dynamical systems. *Discrete Applied Mathematics*, 155 :2403–2413, 2007.

[20] François Robert. Itérations sur des ensembles finis et automates cellulaires contractants. *Linear Algebra and its Applications*, 29 :393–412, 1980.

[21] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65 :386–408, 1958.

[22] René Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42 :563–585, 1973.

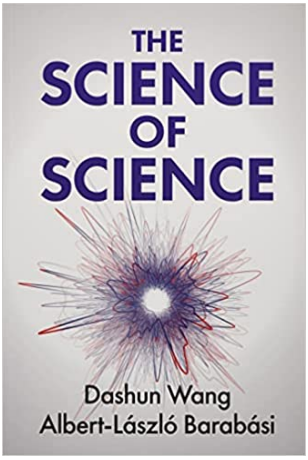
[23] René Thomas. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. In *Numerical methods in the study of critical phenomena*, volume 9, pages 180–193. 1981.

[24] Alan Turing. On Computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society. Second Series*, 45 :230–265, 1937.

The Science of Science

lu par Michel Raynal [1]

The Science of Science
Dashun Wang and Albert-László Barabási, *Cambridge University Press*
ISBN 978-1-108-49266-9, 2021, 303 pages.



Les scientifiques produisent chaque année des millions d'articles, de rapports, de brevets, etc. Dans un tel contexte de données massivement massives, et comme indiqué par les auteurs, la science de la science est un nouveau champ multidisciplinaire dont le but est de comprendre comment la science évolue et ses acteurs (chercheurs) la vivent. Pour cela, Dashun Wang et Albert-László Barabási ont structuré leur livre en quatre parties.

La première partie, intitulée *The science of career* est centrée sur la productivité des chercheurs (pris individuellement) et les différents critères utilisés pour les évaluer, les effets de bord associés (tel que l'effet Matthieu par exemple), le pouvoir prédictif des index utilisés, les processus de relecture des articles (simple aveugle *vs* double aveugle), l'âge des chercheurs par rapport à leur production dans le temps, etc.

1. Université de Rennes IRISA, Inria, CNRS, France.

Constatant que la recherche est de plus en plus menée dans des équipes, la deuxième partie intitulée *The science of collaboration* est consacrée à l'impact de la recherche en équipe sur les résultats. De façon intéressante (avec exemples à l'appui), les auteurs observent par exemple que *large teams develop science, small teams disrupt it*. Cette partie s'intéresse aussi au difficile problème qu'est le « crédit » scientifique attribué à chacun des auteurs (les exemples pris dans des attributions de prix Nobel sont intéressants).

La troisième partie intitulée *The science of impact* est centrée sur la disparité des citations, ce que les citations ne capturent pas, les biais introduits par la presse non-scientifique, l'impact de la formation à l'université sur les prèrequis pour la recherche, etc.

La quatrième partie intitulée *Outlook* pose la question de l'accélération de la science, des biais culturels, et de l'impact de l'intelligence artificielle sur l'ensemble de la démarche scientifique.

Pour conclure, un livre intéressant qui peut aider le lecteur à mieux comprendre les enjeux de la recherche et de son évaluation en lui permettant de former son propre jugement et sa vision critique sur des sujets qui le concernent directement.

Projet ANR (2020-2024) « Hérelles » : un cadre collaboratif unifié pour l'analyse interactive de données temporelles

Pierre Gançarski [1], Antoine Cornuejols [2], Bruno Cremilleux [3], Thi-Bich-Hanh Dao [4], Maguelonne Teisseire [5]

Introduction

Face à la surabondance des données temporelles arrivant de façon quasi-continue, mais aussi à leur complexité et à celle des phénomènes étudiés, définir les informations nécessaires à une analyse de qualité de celles-ci s'avère de plus en plus difficile voire impossible à effectuer a priori. C'est le cas particulièrement lorsque l'on cherche à apprendre un modèle d'attribution de classes à des observations ou à des événements temporels. D'une part, la phase d'étiquetage de l'apprentissage supervisé ne peut plus être assurée par les experts, car trop fastidieuse et chronophage ; de surcroît, dans de nombreux domaines, du fait de la nouveauté de telles données, les champs sémantiques appropriés (principalement les classes à apprendre) peuvent n'avoir pas encore été parfaitement définis. D'autre part, les méthodes aptes à analyser des données temporelles sont elles aussi nombreuses, chacune traitant un type de données de façon privilégiée avec sa propre stratégie d'analyse, et donc ses propres biais.

1. Coordinateur du projet, ICube – université de Strasbourg, gancarski@unistra.fr.
2. AgroParisTech – université Paris-Saclay, antoine.cornuejols@agroparistech.fr.
3. GREYC - université de Normandie, bruno.cremilleux@unicaen.fr.
4. LIFO – université d'Orléans, thi-bich-hanh.dao@univ-orleans.fr.
5. TETIS – INRAE Montpellier, maguelonne.teisseire@irstea.fr.

DOI:10.48556/SIF.1024.20.185 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 185–195

Cette situation rend difficile le choix de la méthode « optimale » et de son paramétrage. Et même si des solutions existent pour combiner de telles méthodes (*boosting, co-training, ensemble learning*, *clustering* collaboratif [2]) et ainsi réduire l'impact de ces choix sur la qualité des résultats, ces approches se limitent majoritairement à la combinaison d'algorithmes partageant le même paradigme (e.g. supervisé ou non supervisé) ce qui limite leur intérêt.

L'objectif du projet HÉRELLES (*Hé*térogénéité des données - Hétérogénéité des méthodes : un cadre collaboratif unifié pour l'analyse interactive de données tempo*relles*) est d'étudier les concepts entrant dans la définition d'une approche fondée sur une collaboration entre des méthodes supervisées et non supervisées, autorisant une découverte progressive de ces informations à travers un cycle d'interactions entre l'expert et le système d'apprentissage. Ainsi, partant d'informations très lacunaires (peu ou pas de classes définies a priori, peu ou pas d'exemples...), l'expert pourra injecter à la volée des connaissances externes non encore exprimées et ce en fonction de l'avancement de l'analyse. Ceci devrait limiter fortement son implication directe et permettre une sémantisation plus aisée. Par ailleurs, cette interaction met en jeu des savoir-faire et des explicitations de connaissances propres au domaine d'application et à l'expert qui sont des biens précieux que le projet ambitionne de capitaliser et analyser pour améliorer les usages futurs du système.

Le consortium formé présente toutes les compétences nécessaires au succès du projet. Ainsi, l'équipe SDC d'ICube (Strasbourg) spécialisée en fouille de données, apporte son expertise en classification collaborative multistratégie de données temporelles guidée par des connaissances ainsi que ses compétences en sciences de l'environnement. Le laboratoire TETIS (Montpellier) spécialisé dans la compréhension des dynamiques des espaces ruraux et des territoires apporte son expertise en géographie, en télédétection et en fouille de données spatio-temporelles. L'équipe CODAG du Greyc (Caen) spécialisée sur les apports mutuels entre fouille de données et approches déclaratives apporte ses compétences sur la prise en compte de préférences explicites et l'élicitation de préférences implicites de l'utilisateur. L'équipe CA du LIFO (Orléans) spécialisée en apprentissage automatique et fouille de données apporte ses compétences en *clustering* sous contraintes et en apprentissage supervisé actif. L'équipe Ekinocs d'AgroParisTech (Paris) particulièrement concernée par l'étude des systèmes complexes multiéchelles spatiales et temporelles apporte son expertise sur les méthodes collaboratives en apprentissage automatique et en fouille de données temporelles.

Hypothèses de recherche

En rupture avec les approches actuelles basées chacune sur un seul paradigme d'analyse, le projet scientifique d'HÉRELLES propose d'étudier et définir une architecture générique (Fig. 1) :

— permettant de faire collaborer des méthodes multiparadigmes travaillant potentiellement sur des données différentes et de définir les conditions optimales de son utilisation [10] ;
— basée sur des mécanismes d'interaction avec l'utilisateur lui offrant la possibilité d'injecter de nouvelles informations et de réduire le fossé sémantique entre les résultats et ses propres intuitions ;
— offrant des méthodes d'extraction et de capitalisation des connaissances directement ou indirectement produites lors du processsus d'analyse ;
— possédant une mise en œuvre effective et une validation dans le cadre de l'analyse de séries temporelles hétérogènes.

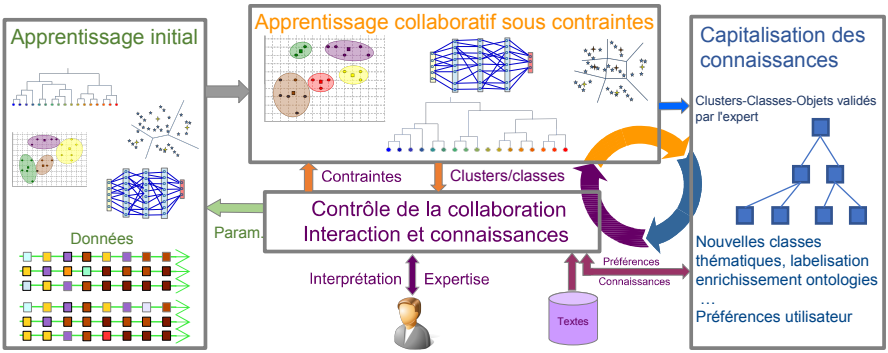


FIGURE 1. Une approche interactive pour la découverte de connaissances

Une approche collaborative. Concrètement, il s'agira de concevoir une approche autorisant la collaboration entre des algorithmes d'apprentissage différents à travers des transferts d'informations extraites à partir de données temporelles complémentaires. Par exemple, les clusters seront utilisés pour étiqueter des données ou comme classes thématiques potentielles, les données d'apprentissage serviront à la validation des clusters ou pour générer des contraintes, des informations sur des clusters pourront être partagées par toutes les méthodes... Cette collaboration sera contrôlée par des critères internes (convergence des résultats, adéquation avec des connaissances a priori...) et par l'expert qui interviendra en évaluant des résultats intermédiaires et en injectant de nouvelles informations (contraintes entre des objets, labélisation d'objets...).

Une approche active. L'expérience montre que demander à l'expert de fournir a priori toutes les informations nécessaires au processus de traitement des données n'est pas réaliste. Rendre incrémentales les méthodes d'apprentissage permettra à

l'utilisateur de guider le processus d'analyse en injectant de nouvelles données mais aussi, et surtout, de nouvelles connaissances. Cependant, devant la complexité ou la richesse des résultats, l'expert est souvent démuni et ne peut généralement pas proposer des modifications de ceux-ci ou l'ajout de données ou d'informations qui soient réellement pertinents. L'apprentissage actif permet de guider le processus d'analyse de données grâce à des phases de questions et de réponses dans lesquelles l'algorithme soumet à l'expert des propositions à infirmer ou confirmer. En général, ces questions découlent de l'analyse du résultat courant et se concentrent sur les objets difficiles à classer ou à affecter à un cluster. Dans ce projet, nous proposons que cette approche soit étendue à la découverte d'informations de plus haut niveau comme par exemple, les classes manquantes (existantes ou à créer dans la nomenclature) ou les liens sémantiques entre ces classes et les clusters en définissant des mécanismes d'analyse des résultats au regard de connaissances thématiques (ontologies, nomenclature...) ou de données complémentaires (textes, page web...). Parallèlement à l'étude de ces extensions, une part importante du projet sera consacrée au développement d'une interface d'interaction à la fois conviviale et efficace permettant de réduire l'effort cognitif de l'expert pour d'une part, comprendre les résultats et les questions posées par le système et d'autre part, y répondre de façon « optimale ».

Une approche générique. La méthode proposée sera générique et ne dépendra pas des domaines potentiels d'application. Afin d'en valider l'opérabilité, nous nous intéresserons à la compréhension de phénomènes complexes de notre environnement (artificialisation des sols, urbanisation, construction d'infrastructures…) majoritairement via des données hétérogènes majoritairement satellitaires et textuelles domaines dans lequels l'expertise de TETIS et d'ICube est avérée. Pour la mise en œuvre concrète, nous nous appuierons sur la plateforme existante FODOMUST [6] développée par l'équipe SDC. Celle-ci sera étendue pour d'une part, intégrer toutes les avancées méthodologiques et algorithmiques et d'autre part, compléter l'interface MULTICUBE permettant une utilisation distante via un accès au serveur hébergé par l'université de Strasbourg.

Application à l'observation de la Terre

Notre volonté est d'aller plus loin qu'une simple preuve de concept. Pour cela, la plateforme FODOMUST sera le support d'une validation méthodologique via des mécanismes génériques et d'une validation thématique via des interfaces et développements dédiés à l'observation de la Terre. En effet, pour appréhender, observer, comprendre les dynamiques des paysages, les experts en sciences de la Terre et de l'environnement disposent de volumes considérables et toujours croissants de données d'observation satellitaires arrivant de façon massive en flux quasi continu. Avec le lancement et la mise en production des satellites européens de la constellation

Sentinel, ou franco-israélien Venμs, ces images de la Terre sont acquises à une fréquence jamais encore atteinte. De plus, elles sont de plus en plus souvent mises librement à la disposition des scientifiques, des décideurs ou des utilisateurs du domaine. Néanmoins, utiliser de façon optimale ces données reste un défi. En effet, peu d'utilisateurs sont capables d'exploiter tout le potentiel offert par des séries temporelles d'images satellitaires sans une aide leur permettant de relier les informations qu'elles fournissent entre elles ou à celles issues d'autres sources et surtout à leurs propres connaissances. Appliquer les propositions faites par HÉRELLES permettra de confronter l'approche à la réalité et ainsi de la valider concrètement, mais aussi de profiter des retours d'expériences afin d'en améliorer l'interface et le mode de fonctionnement.

Un exemple de cas d'étude type. L'artificialisation des sols est un enjeu important, notamment pour la métropole de Montpellier Méditerranée (3M) car cette agglomération se développe de façon très active. La croissance démographique s'y illustre par un accroissement des surfaces bâties et des infrastructures aux dépens des zones naturelles ou agricoles. Pour appréhender et comprendre les déterminants de cette artificialisation, que la seule démographie ne suffit pas à cerner, le cas d'étude se fondera sur des analyses spatio-temporelles à travers des jeux de données hétérogènes d'imagerie satellitaire associés à d'autres sources de données (textes, bases de données, etc.). L'objectif est de fournir des clés qui permettent à la collectivité d'accueillir les nouveaux citoyens tout en maîtrisant son artificialisation et en étant responsable face aux objectifs de développement durable. Ce cas d'étude s'inscrit dans les quatre hypothèses formulées dans le cadre du projet HÉRELLES. Une méthodologie collaborative multiparadigme permettra de répondre aux analyses demandées par les acteurs sur le territoire car celles-ci sont variées et fortement dépendantes des données et connaissances disponibles. Une approche incrémentale et interactive se justifie pleinement pour une co-construction des connaissances des dynamiques territoriales. La sémantisation est indispensable à l'appropriation et la capitalisation des résultats obtenus. Enfin, la dimension temporelle est un élément incontournable de l'analyse des dynamiques.

Programme scientifique

Outre une tâche dédiée à la coordination du projet, nous avons structuré le programme scientifique en quatre tâches, chacune étant portée par deux partenaires.

Collaboration multiparadigme (T1). L'objectif est de proposer une nouvelle architecture opérationnelle et un cadre théorique pour faire collaborer des techniques d'apprentissage supervisé et non supervisé dans une procédure itérative, d'étudier les conditions de la mise en œuvre d'une telle architecture et enfin de définir des mécanismes et conditions d'optimisation du processus. Pour cela, dans un premier temps, nous nous appuierons sur les principaux concepts mis en œuvre dans la méthode de

clustering collaboratif SAMARAH [8]. Nous y reprendrons l'hypothèse de base à savoir que la qualité d'une solution globale optimale ou satisfaisante pour l'expert est fortement corrélée à la qualité individuelle de chacun des résultats locaux produits par les méthodes impliquées dans la collaboration, leur similarité ainsi que le respect des connaissances et contraintes imposées par l'expert (labels initiaux des objets, contraintes de clustering...). Nous nous inspirons du processus de raffinement des résultats locaux basé sur des interactions (détection de divergences et tentative de résolution de celles-ci) entre les méthodes prises deux à deux. Étendre cette approche vers une architecture multiparadigme autorisant la collaboration entre des méthodes supervisées et non supervisés nécessite de lever quatre verrous :

— quelles mesures de compatibilité ou de désaccord entre méthodes supervisées, non supervisées et mixtes faut-il définir ?
— quelle est la stratégie optimale permettant aux agents impliqués de modifier localement leurs résultats en vue de maximiser la compatibilité ou de réduire ces désaccords ?
— comment trouver un formalisme commun de communication entre les différentes méthodes pour leurs échanges de données et de résultats ?
— quelle stratégie globale innovante doit être mise en œuvre afin de garantir une convergence de l'ensemble de classifieurs vers une solution satisfaisante pour l'expert ?

Dans un deuxième temps, si la durée du projet nous le permet, nous chercherons à optimiser la stratégie mise en œuvre en tentant de caractériser l'effet de la diversité et de la décorrélation des informations apportées par les méthodes [10] afin de garantir la convergence des processus et la prévention des collaborations négatives pouvant amenant une dégradation du résultat global ou des résultats individuels des méthodes.

Incrémentalité et interactions (T2). L'objectif est, d'une part, d'intégrer dans des méthodes d'apprentissage (principalement sous contraintes [1, 7, 3]) des mécanismes efficaces leur permettant d'exploiter de nouvelles données ou informations (nouvelles classes, labélisations, contraintes...) fournies par l'expert à la volée et d'autre part, de concevoir une méthode d'interaction active en proposant des mécanismes de sélection des informations ou questions à soumettre à l'expert et d'optimisation des retours de celui-ci.

Pour cela, dans un premier temps, nous rendrons incrémentales les méthodes d'apprentissage tout en optimisant la prise en compte des nouvelles informations. En effet, si l'expert ne voit pas l'amélioration notable de la solution grâce à son aide, il perdra rapidement confiance dans le système. Paradoxalement, il est préférable que tout nouveau résultat ne nécessite pas un effort cognitif important de la part de l'expert, du fait de la perturbation consécutive à cette information (autrement dit, le

nouveau résultat ne devra pas trop s'éloigner du précédent) au risque de le dérouter. Nous ferons en sorte de lever trois verrous principaux :

— comment tirer parti des indications venant de l'expert afin que celles apportant le plus d'information ou considérées comme les plus importantes par l'expert soient satisfaites en priorité ?
— comment limiter les perturbations de la solution courante issue de la collaboration en cours ?
— quels mécanismes faut-il mettre en place pour détecter et résoudre les incohérences potentielles induites par les contraintes ajoutées à la volée ?

Dans un deuxième temps, nous nous intéresserons à une méthode d'apprentissage actif générant automatiquement les questions sur des classes, labels, contraintes... à soumettre à l'expert. Les informations validées seront alors transmises au processus collaboratif. Il s'agit ici, là encore, de lever trois verrous :

— les différences entre les résultats dues à l'hétérogénéité des méthodes peuvent-elles être exploitées ?
— peut-on se baser sur un historique des retours des experts pour tenter d'identifier les contraintes les plus informatives et évaluer leur impact sur le modèle et ainsi en limiter les perturbations ?
— comment présenter les résultats et ces propositions afin d'aider l'expert à valider ou infirmer les propositions.

Il apparaît de façon évidente que l'interface nécessaire à cette interaction est cruciale et dépend fortement des domaines d'application visés. Sa conception sera étudiée dans la tâche T4.

Capitalisation des connaissances (T3). L'objectif est de concevoir des mécanismes originaux favorisant la sémantisation des résultats (nommage des entités et des classes, labélisation ou étiquetage d'objets ou de clusters...) et des profils d'évolution détectés, de proposer de nouveaux moyens d'identification et d'apprentissage des besoins et préférences de l'utilisateur et de mise en place de mécanismes de génération à partir de connaissances du domaine (ontologies, textes...) de contraintes exploitables par le processus collaboratif. Pour cela, dans un premier temps, nous nous concentrerons sur des mécanismes permettant d'associer une sémantique à des objets d'intérêt (éléments géographiques, classes ou clusters...) en s'appuyant sur une mise en correspondance avec des éléments textuels extraits de données exogènes (journaux, textes réglementaires...) [4] analysés selon leurs dimensions spatiales, temporelles et thématiques [9, 7, 5]. Il s'agit toujours de lever trois verrous :

— quel mécanisme d'extraction des éléments textuels exogènes potentiellement intégrables dans le processus de sémantisation doit-on définir ?
— comment ancrer spatialement et temporellement ces éléments aux résultats à partir des métadonnées disponibles sur ces derniers et des descripteurs spatio-temporels issus de textes ?

— quand et comment faire intervenir l'expert dans la définition et le choix des descripteurs caractérisant les clusterings et contribuant à leur sémantisation.

Dans un deuxième temps, si la durée du projet nous le permet, nous nous efforcerons de concevoir une méthodologie pour éliciter les préférences exprimant l'intérêt de l'utilisateur. En effet si l'utilisateur a une idée de ce qu'il cherche, il a souvent du mal à exprimer ses besoins et à les traduire en des mécanismes opérationnels par les méthodes de clustering.

Plateforme FODOMUST *et validation (T4).* Il s'agit ici de valider les avancées méthodologiques et thématiques dans le domaine particulier de l'observation de la Terre, de les pérenniser dans une plateforme opérationnelle munie d'une interface conviviale permettant d'interagir avec le processus. Or, évaluer un projet de recherche exploratoire, bien que présentant de fortes connexions avec des domaines applicatifs, est par nature difficile. Les objectifs visés par HÉRELLES sont méthodologiques, voire théoriques, et non totalement garantis et de fait difficilement quantifiables. Pour les apports méthodologiques, nous nous appuierons principalement sur des validations statistiques via les critères classiques d'évaluation des méthodes d'apprentissage. Nous les compléterons par des critères propres aux approches collaboratives telles que la convergence des agents ou l'influence de la diversité sur la qualité des solutions. Pareillement, nous nous intéresserons à une validation thématique pour laquelle, il n'existe malheureusement pour l'instant que très peu de données de référence. Or, en l'absence de vérité terrain, voire de connaissances du domaine formalisées, il a été démontré que les critères classiques peuvent être faillibles et ne sont que peu utiles dans ce contexte. Ainsi, la validation thématique ne pourra se faire qu'à dire d'expert dans un premier temps. Néanmoins, nous essaierons, dès que cela est possible et si cela présente un intérêt scientifique, de rattacher ces résultats à des données de référence monodates. Nous nous efforcerons de faire évaluer par les communautés concernées les apports de la méthode dans leur domaine d'expertise. Nous nous intéresserons principalement aux capacités de notre méthode à proposer des interactions les plus efficaces possibles tout en imitant l'implication de l'expert. Il s'agira aussi, tout au long du projet, d'apprécier la facilité d'usage et d'analyser la performance des méthodes via des enquêtes, ateliers dédiées, hackathons… Une definition participative des critères d'évaluation ou de validation associée à une coconception de la méthode nous semble le moyen le plus sûr d'atteindre cet objectif de validation.

Une architecture innovante pour une approche innovante

La première année de projet, en plus d'avancées ponctuelles mais néanmoins très

intéressantes dans chacune des tâches, nous a permis de formaliser le processus global d'analyse de séries temporelles (d'images satellitaires dans notre cas) et d'extraction de connaissances à partir de données exogènes. La figure 2 présente visuellement ce processus et son lien avec les différentes tâches (T_i). L'expert est au cœur du processus. En validant (ou non) les propositions faites par le système ou en injectant de nouvelles connaissances exogènes, il influe fortement sur le processus. Les premières expérimentations faites dans le cadre du suivi de coupes forestières par l'approche collaborative incrémentale semblent valider notre hypothèse en donnant des résultats très encourageants.

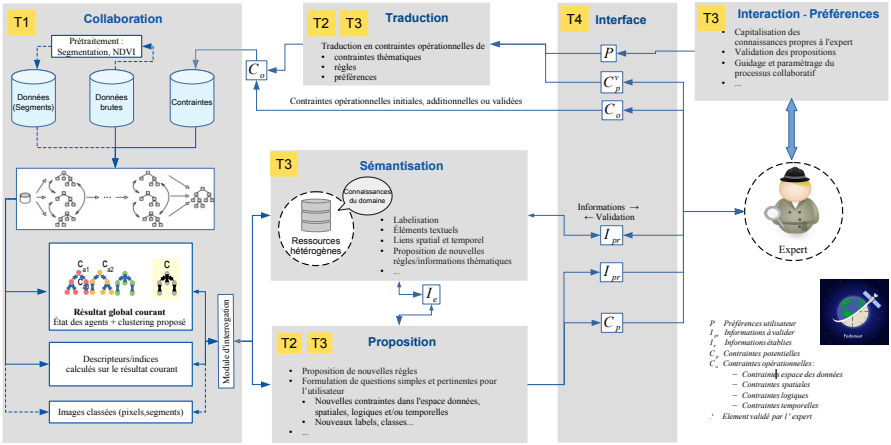


FIGURE 2. Architecture générale et découpage en tâches.

Impacts et retombées du projet

Avancées méthodologiques. Ce projet induira des retombées scientifiques multiples que nous tâcherons de diffuser massivement. On peut citer, au niveau global, la définition, l'intégration et l'optimisation d'une nouvelle méthode collaborative multi-paradigme dans une plateforme ouverte à des usages dans des domaines applicatifs variés ; le développement d'outils et de méthodes d'interaction entre le système et l'expert l'autorisant à intervenir sur le processus incrémental afin de limiter fortement ses interventions ou encore l'étude de l'applicabilité de contraintes thématiques à des processus non supervisés de fouille de données spatio-temporelles. L'usage et la validation thématique en observation de la Terre depuis cette plateforme devraient

permettre la découverte incrémentale de nouvelles classes d'évolution, voire de taxonomies nouvelles, indispensables à la compréhension des dynamiques des paysages. Nous veillerons à les présenter sous une forme diffusable et compréhensible par les différents acteurs et décideurs dans ce domaine. Concrètement, le projet débouchera sur une solution logicielle ouverte et libre de droit pérennisée via la plateforme FODOMUST ; ce qui devrait permettre un essor de ces approches.

Jeux de données et résultats. En plus de la valorisation des résultats scientifiques par des publications et communications, nous diffuserons des jeux de données de référence (*benchmark*) vers la communauté scientifique, action devenue à la fois classique et obligatoire dans le domaine de la fouille de données et de l'apprentissage. Ainsi, à l'issue du projet, les bases de données initiales, de même que celles constituées à partir des résultats obtenus, seront ouvertes et librement utilisables.

Dynamique des paysages. Enfin nous encouragerons les publications interdisciplinaires, en particulier autour de la géomatique et de la télédétection voire du TALN (revues RIG, IJGIS et conférences SAGEO, AGILE et SIG Spatial...). De même, des communications auront lieu, par exemple, dans le cadre des journées et ateliers des GdR MAGIS ou dans le cadre d'un atelier organisé dans la communauté d'agglomération de Montpellier, en collaboration avec le syndicat mixte du bassin de Thau (SMBT), ainsi qu'avec les participants à THEIA (en particulier ceux du CES Détection de changement).

Conclusion

Le projet HÉRELLES est ambitieux par sa volonté d'apporter des solutions dans l'analyse de données temporelles pour lesquelles l'expertise est encore incomplète et insuffisante pour une approche supervisée. En premier lieu, il s'adresse à des experts de domaines pour lesquels il n'existe pas encore de nomenclature et de bases d'apprentissage complètement définies ou opérables. Pour pallier ce manque, tout en limitant l'investissement de l'expert, la solution proposée se fonde sur une découverte progressive pendant l'exploitation et l'analyse des données à travers un cycle d'interactions entre l'expert et le système d'apprentissage. Cette interaction vise à réduire le fossé entre les résultats produits par les algorithmes et les intuitions thématiques de l'expert et rendre les résultats plus compréhensibles pour celui-ci, donc plus aisés à associer à une sémantique du domaine d'application. De plus, afin de limiter fortement les choix initiaux (données à analyser, données d'apprentissages, méthodes et paramètres) à toute analyse, il s'appuie sur les concepts mis en œuvre dans l'approche de clustering collaboratif incrémental SAMARAH qui a fait ses preuves dans de nombreux domaines. Son principal domaine d'expérimentation et de validation est celui des sciences de la Terre et de l'environnement. En effet si l'usage de l'imagerie satellitaire y est fortement répandu en analyse monodate, il est encore très

limité, dès qu'il s'agit d'analyses temporelles. En effet, la mise à disposition de données temporelles à haute fréquence d'acquisition est encore relativement récente. De fait, excepté quelques domaines tel que, par exemple, l'agriculture, il n'existe pas de telles nomenclatures et donc de données d'apprentissage opérables. Concevoir une plateforme complète et totalement utilisable en 4 ans est un défi. Néanmoins, nous sommes confiants sur notre capacité à produire des résultats scientifiques et des prototypes qui pourront servir de base à des travaux plus approfondis tant théoriquement qu'en termes de logiciels de qualité. Nous invitons toute personne intéressée par ces perspectives à nous contacter.

Remerciements

Ces travaux sont soutenus par l'Agence nationale de la recherche dans le cadre du projet Hérelles (ANR-20-C23-0022).

Références

[1] Aymeric Beauchamp, Thi-Bich-Hanh Dao, Samir Loudni, and Christel Vrain. Intégration incrémentale de contraintes pour le clustering avec la programmation par contraintes. In *PFIA - Journées Francophones de Programmation par Contraintes, Saint-Étienne*, 2022.

[2] Antoine Cornuéjols, Cédric Wemmert, Pierre Gançarski, and Younes Bennani. Collaborative clustering : Why, when, what and how. *Information Fusion*, 2018, volume = 39, pages = 81–95,.

[3] Thi-Bich-Hanh Dao, Khanh-Chuong Duong, and Christel Vrain. Constrained clustering by constraint programmin. *Artificial Intelligence*, 244 :70–94, 2017.

[4] Roche Mathieu et al. Animitex : Analyse d'images fondée sur des informations textuelles. *Revue des Sciences et Technologies de l'Information*, 19(13) :163–167, 2014.

[5] Jérôme Fize, Mathieu Roche, and Maguelonne Teisseire. Could spatial features help the matching of textual data ? *Intelligent Data Analysis*, 24(5) :1043–1064, 2020.

[6] SDC FoDoMuST. Version interface déportée - client léger. `https://sdc.icube.unistra.fr/en/index.php?title=FODOMUST`, Août 2022.

[7] Pierre Gançarski, Thi-Bich-Hanh Dao, Cremilleux Crémilleux, Germain Forestier, and Thomas Lampert. *A Guided Tour of AI Research*, chapter Constrained Clustering : Current and New Trends. Springers, 2020.

[8] Pierre Gançarski and Cédric Wemmert. Collaborative multi-step mono-level multi-strategy classification. *Multimedia Tools and Applications*, 35(1) :1–27, 2007.

[9] Urcel Kalenga-Tshingomba, Magali Jouven, Lucile Sautot, Imad Shaqura, and Maguelonne Teisseire. Cartographie des surfaces pastorales à l'aide des données sentinel 2 l3a et des données ouvertes : promesses et réalités. *Revue Internationale de Géomatique*, À paraitre, 2022.

[10] Parisa Rastin, Barasab Matei, Guénaël Cabanes, and Younès Grozavu, Nistor et Bennani. Impact of learners' quality and diversity in collaborative clustering. *Journal of Artificial Intelligence and Soft Computing Research*, 9(12) :149–165, 2019.

Le problème des trois cartes

Jean-Paul Delahaye [1]

La rubrique « Récréation informatique » propose une petite énigme algorithmique ou sur un thème de mathématiques discrètes susceptible d'intéresser un lecteur de 1024. La solution est donnée dans le numéro suivant.

Rappel et solution du problème précédent

Étonnantes puissances de 2

Comme tout informaticien, vous êtes fascinés par les puissances de 2 : 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048... Si on observe le dernier chiffre (chiffre des unités), il est immédiat de constater et de démontrer que la série '2', '4', '8', '6' y est répétée indéfiniment : le dernier chiffre est donc dans 25 % des cas un '2', dans 25 % des cas un '4', dans 25 % des cas un '8' et dans 25 % des cas un '6'. Si on observe les deux derniers chiffres, on remarque (et on démontre sans mal encore) que la séquence '04', '08', '16', '32', '64', '28', '56', '12', '24', '48', '96', '92', '84', '68', '36', '72', '44', '88', '76', '52' revient périodiquement. On en déduit que le deuxième chiffre à partir de la droite (chiffre des dizaines) est périodiquement '0', '0', '1', '3', '6', '2', '5', '1', '2', '4', '9', '9', '8', '6', '3', '7', '4', '8', '7', '5'. Cette séquence périodique de longueur 20 comporte chaque chiffre deux fois exactement, ce qui a pour conséquence que le '0' sera le deuxième chiffre de 2^k à partir de le droite (chiffre des dizaines) dans 10 % des cas, et que c'est vrai aussi pour le '1', le

1. Professeur émérite, université de Lille, campus scientifique, CRIStAL UMR CNRS, 9189 Centre de recherche en informatique signal et automatique de Lille, bâtiment ESPRIT, 59655, Villeneuve d'Ascq Cedex France. E-mail : jean-paul.delahaye@univ-lille.fr.

DOI:10.48556/SIF.1024.20.197 **1024** – Bulletin de la SiF, numéro 20, Novembre 2022, pp. 197–200

'2', etc. Les chiffres en deuxième position à partir de la droite dans 2^k sont présents chacun avec une fréquence de 10 %. L'étude du chiffre en troisième position à partir de la droite (chiffre des centaines) conduit à la même conclusion, et il en va de même aussi pour le quatrième, etc. La propriété n'est pas facile à démontrer mais elle est vraie et avec un programme, on la vérifie expérimentalement [2].

Il est donc vrai que sauf pour le chiffre des unités, les chiffres des puissances de 2 placés en position n à partir de la droite ont chacun une fréquence d'apparition de 10 % et cela, quel que soit $n > 1$.

Considérons maintenant le second chiffre à gauche des 2^n qui peut être 0, 1, ..., 9. Une étude numérique montre, cette fois, que les dix chiffres n'ont pas la même fréquence d'apparition :

0	1	2	3	4	5	6	7	8	9
0.120	0.114	0.109	0.104	0.100	0.097	0.093	0.090	0.087	0.084

FIGURE 1. Fréquence d'apparition du second chiffre à gauche de 2^n.

Parfois c'est plus que 0.1, parfois moins. Nous n'avons pas considéré le premier chiffre à gauche car il ne peut pas être '0' et donc, nous ne pouvions pas espérer avoir une fréquence d'apparition égale à 0.1.

N'est-il pas paradoxal que l'égalité des fréquences des 10 chiffres qu'on constate et qu'on démontre pour le chiffre des dizaines, pour le chiffres des centaines, etc. n'entraîne pas l'égalité des fréquences pour ce second chiffre à gauche ?

Comment l'expliquer ?

SOLUTION.

Le résultat concernant le second chiffre à gauche provient de la loi de Benford qui a été démontrée pour la suite des puissances de deux [1].

En base de numération b, la probabilité d'apparition du chiffre c en tête d'une puissance de deux est donc $\log_b(1+\frac{1}{c})$. Un intéressant résultat indique que la probabilité qu'un nombre appartenant à une série de nombres vérifiant la loi de Benford commence par un entier quelconque c écrit en base b, par exemple $c = 314$, est aussi $\log_b(1+\frac{1}{c})$.

La probabilité pour que le second chiffre à gauche en base 10 de 2^n soit un '0' est la probabilité qu'il commence par '10', ou '20', ..., ou '90'. C'est donc $\log_{10}(1+\frac{1}{10})+\log_{10}(1+\frac{1}{20})+\cdots+\log_{10}(1+\frac{1}{90}) \approx 0,1196792685$. De même, la probabilité pour que le second chiffre à gauche en base 10 de 2^n soit un '1' est la probabilité qu'il commence par '11', ou '21', ..., ou '91'. C'est donc $\log_{10}(1+\frac{1}{11})+\log_{10}(1+\frac{1}{21})+\cdots+\log_{10}(1+\frac{1}{91}) \approx 0,1138901039$. Et ainsi de suite...

Les calculs provenant des résultats (démontrés) de la théorie coïncident sans surprise avec ceux obtenus par programme donnés dans l'énoncé. On aura donc beau

mener les calculs en faisant tourner très longtemps son ordinateur, on ne trouvera jamais des fréquences égales pour les dix chiffres possibles en seconde position dans l'écriture décimale de 2^n.

Il n'y a cependant aucune contradiction entre ce qui se produit pour le second chiffre de 2^n et ce qui se passe pour les chiffres comptés à partir de la droite (chiffres des dizaines, chiffres des centaines, etc.). La chose est un peu étonnante et provient de la remarque suivante. Le second chiffre à gauche n'est pas toujours au même rang quand on compte à partir de la droite. Après 1, 2, 4, 8 qui n'ont qu'un chiffre, dans un premier temps, pour 16, 32, 64 ; le second chiffre à partir de la gauche est le même que le premier chiffre à partir de la droite. Mais pour 128, 256, 512, le second chiffre à partir de la gauche est devenu le second chiffre à partir de la droite. Un peu plus loin cela change encore, et cela change tout le temps car la longueur de l'écriture décimale de 2^k augmente indéfiniment. Cela rend compatibles les diverses affirmations : l'égale répartition dans l'écriture décimale de 2^n pour les chiffres des dizaines, pour les chiffres des centaines, etc. n'oblige pas à une égale répartition pour le second chiffre significatif. Si vous avez des doutes, la seule solution est de contrôler par programme tout ce qui a été affirmé.

Nouveau problème

LE PROBLÈME DES TROIS CARTES

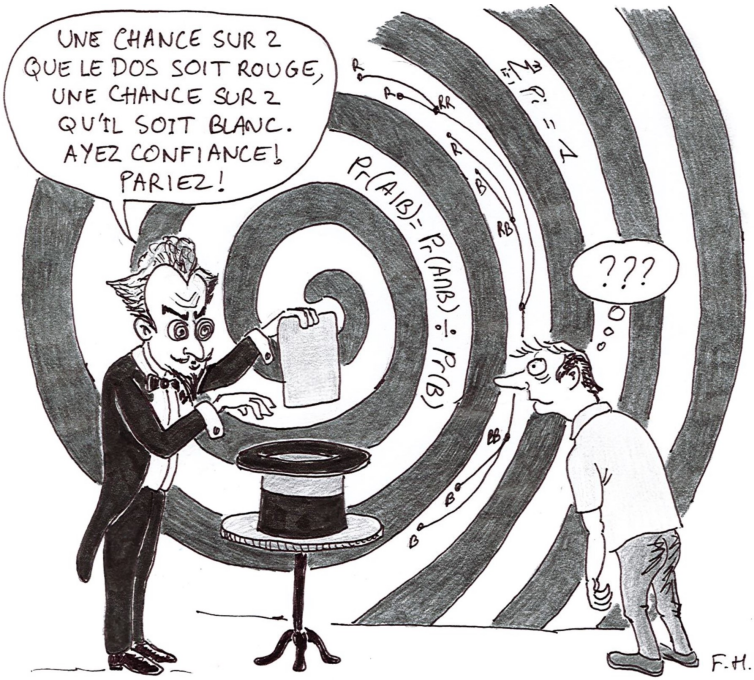
Le présentateur met trois cartes dans un chapeau. La première carte est rouge sur ses deux faces (*RR*), la seconde blanche sur ses deux faces (*BB*), la troisième rouge d'un côté et blanche de l'autre (*RB*).

Le présentateur demande à une personne qui a les yeux bandés de prendre une carte du chapeau, de la sortir en n'en montrant qu'une face au hasard. La face montrée est rouge. On s'intéresse alors à la couleur non visible de la carte choisie.

Le présentateur vous propose un pari sur la couleur de la face non visible de la carte choisie : il vous donnera cinq euros si la couleur choisie est blanche, vous lui donnerez quatre euros si elle est rouge. Est-il dans votre intérêt de parier ?

Vous vous dites : « La carte choisie est l'une des cartes *RR* ou *RB*. Si c'est la première, la face cachée est rouge, si c'est la seconde la face cachée est blanche. La carte choisie a été tirée au hasard, il y a donc une chance sur deux que ce soit l'une ou l'autre, et donc que je gagne. Le pari proposé m'est donc favorable. J'accepte. »

Vous avez tort d'accepter. Pourquoi ?



Envoyez vos réponses à jean-paul.delahaye@univ-lille.fr. Les noms des premiers lecteurs à me donner la bonne réponse (et à la justifier) seront mentionnés dans le prochain numéro de 1024.

Références

[1] Arnold, V.I. and Avez, A. *Ergodic Problems of Classical Mechanics.* Mathematical physics monograph series, ed. Benjamin, 1968.

[2] Michael Fuchs Digital Expansion of Exponential Sequences J. Theor. Nombres Bordeaux, 14 :2, 477-487, 2002.

ADHÉRER À LA SIF

La Société informatique de France est un espace de réflexion, de concertation sur les enjeux de l'informatique, mais aussi un espace d'actions, basé sur le travail de la communauté, qui vise à rassembler toutes celles et tous ceux pour qui faire progresser l'informatique est un métier ou une passion : enseignants, chercheurs, ingénieurs, industriels, consultants et étudiants.

Les adhésions sont valables 12 mois à compter de la date d'adhésion.

Personnes physiques

Tarif plein : 30 €
Tarif réduit : 15 €

- Membre d'un adhérent institutionnel de la SIF.
- CDD, CDI depuis moins de 2 ans, retraité.
- Membre d'une association partenaire, ou de l'ACM.

Gratuit : étudiants, doctorants et post-doctorants.
La SIF vous offre la possibilité d'effectuer le règlement de la cotisation directement en ligne.

Partenaires (Personnes morales)

Les associations partenaires, membres du Conseil des associations de la SIF, ne paient pas de cotisation. Les institutions telles que laboratoires, unités d'enseignement, ou entreprises, peuvent adhérer en tant que telles à la SIF. Il n'existe pas de tarif spécifique pour les adhérents institutionnels : en fonction de leur taille, de leur secteur d'activité, l'importance de l'effort ne se mesure pas de la même façon.

La SIF propose cinq niveaux de cotisation. Pour vous aider dans votre choix, vous trouverez ci-dessous une indication du tarif en fonction de la taille de l'adhérent institutionnel :

- Tarif 1 : 250 € (moins de 50 personnes)
- Tarif 2 : 500 € (de 50 à 100 personnes)
- Tarif 3 : 1000 € (de 100 à 150 personnes)
- Tarif 4 : 1500 € (de 150 à 200 personnes)
- Tarif 5 : 2000 € (au delà de 200 personnes)

Pour adhérer, vous devez contacter notre trésorier. Pour toute question, ne pas hésiter à contacter notre secrétariat.

Plus d'informations sur notre site internet :

`https://www.societe-informatique-de-france.fr`

1024 Bulletin de la Société informatique de France

Société informatique de France

Institut H. Poincaré · 11 rue Pierre & Marie Curie · 75231 PARIS Cedex 05