



**HAL**  
open science

# RECONSTRUCTION ET MODÉLISATION DES DYNAMIQUES SOCIALES ET DE L'ÉVOLUTION CULTURELLE

David Chavalarias

► **To cite this version:**

David Chavalarias. RECONSTRUCTION ET MODÉLISATION DES DYNAMIQUES SOCIALES ET DE L'ÉVOLUTION CULTURELLE : Le tournant des Sciences Humaines et Sociales du XXIème siècle. Ordinateur et société [cs.CY]. EHESS-Paris, 2016. tel-01394843

**HAL Id: tel-01394843**

**<https://hal.science/tel-01394843>**

Submitted on 10 Nov 2016

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

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



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0  
International License

# RECONSTRUCTION ET MODÉLISATION DES DYNAMIQUES SOCIALES ET DE L'ÉVOLUTION CULTURELLE

Le tournant des Sciences Humaines et Sociales  
du XXIème siècle

**David Chavalarias**

Chercheur au Centre d'Analyse et de Mathématique Sociales (CAMS)  
(UMR 8457)

Directeur de l'Institut des Systèmes Complexes de Paris Île-de-France  
(UPS3611)

*Mémoire en vue de l'habilitation à diriger des recherches*

*Soutenue le 23 mai 2016*

## **Membres du Jury**

- Jean-Pierre Nadal, Directeur de Recherche CNRS, ENS/EHESS UMR8557 (garant),
- Serge Abiteboul, Directeur de Recherche Inria et professeur affilié à l'ENS Cachan, Etablissement de rattachement : Inria (examinateur),
- Madeleine Akrich, Directrice de recherche, Centre de sociologie de l'innovation, Mines ParisTech (présidente du jury),
- Francis Chateauraynaud, Directeur d'Études en Sociologie à l'EHESS (examinateur),
- Alan Kirman, Professeur émérite de Sciences Économiques à l'Université d'Aix-Marseille III et Directeur d'Études à l'ÉHESS (rapporteur),
- Emmanuel Lazega, Professeur des Universités à Sciences-Po en Sociologie (rapporteur).

# Table des matières

<b>I</b>	<b>Parcours de recherche</b>	<b>4</b>
1	Régularités, lois et modèles en SHS . . . . .	5
1.1	Qu'y a-t-il à modéliser ? . . . . .	5
1.1.1	Les enseignements de la physique . . . . .	5
1.1.2	Les enseignements de la biologie . . . . .	8
1.2	La fonction du modèle . . . . .	11
1.3	L'étude des modèles . . . . .	17
1.4	Reconstruction théorique, reconstruction phénoménologique . . . . .	18
1.4.1	<i>Reconstruction théorique</i> . . . . .	19
1.4.2	<i>Reconstruction phénoménologique</i> . . . . .	19
1.4.3	Articulation des opérations de reconstruction . . . . .	21
2	Principaux objets de recherche . . . . .	23
2.1	Le triptyque de la modélisation en sciences humaines et sociales . . . . .	24
2.2	Modèles et méga-données . . . . .	25
2.3	A propos de l'usage des termes <i>évolution culturelle et cognition sociale</i> . . . . .	29
2.3.1	Évolution culturelle . . . . .	29
2.3.2	Cognition sociale . . . . .	29
2.4	Résumé . . . . .	30
3	Reconstruction des dynamiques sociales à l'ère des méga-données . . . . .	31
3.1	La thèse de Popper est-elle réfutable ? . . . . .	33
3.1.1	La philosophie des sciences comme objet de modélisation . . . . .	33
3.1.2	Nobel, le jeu de la découverte collective. . . . .	33
3.2	Reconstruction de la dynamique des sciences . . . . .	37
3.2.1	Les analyses par co-citations et par co-termes . . . . .	38
3.2.2	Phylométrie de sciences . . . . .	41
3.2.3	Une méthodologie transversale . . . . .	42
3.2.4	Vers une analyse morphologique de l'évolution des sciences . . . . .	44
3.3	Qualité des théories et évolution des normes de publication . . . . .	49
3.4	Le Tweetoscope climatique : explorer les dynamiques culturelles autour d'un enjeu de société . . . . .	51
3.5	Enjeux de l'analyse quantitative de l'évolution des sciences et de l'évolution culturelle . . . . .	54
3.5.1	Qualification multi-échelles de la morphologie des phylométries . . . . .	54
3.5.2	Temporalités et circulation de l'attention sociale . . . . .	54
3.5.3	Réseaux phylométriques et dynamiques sociales . . . . .	55
3.5.4	Réseaux phylométriques multi-partites . . . . .	56

3.5.5	Réseaux phylométriques et modélisation de l'évolution des sciences . . . . .	57
3.6	Feuille de route technologique pour l'étude phylométrique des sciences . . . . .	57
3.6.1	Passage à l'échelle des algorithmes de reconstruction . . . . .	57
3.6.2	Crowdsourcing et visualisation interactive . . . . .	57
3.7	Résumé . . . . .	59
4	Modéliser l'évolution culturelle . . . . .	61
4.1	Modélisation de la formation endogène des réseaux . . . . .	62
4.2	La cognition sociale et la question de la différenciation . . . . .	64
4.2.1	Entre individualisme méthodologique et holisme . . . . .	64
4.2.2	L'imitation, trait d'union entre l'individu et le collectif . . . . .	65
4.2.3	Désir mimétique et imitation rationnelle . . . . .	66
4.2.4	Repenser l'imitation . . . . .	69
4.3	Éthique des TIC . . . . .	74
5	Conclusions . . . . .	79
	Références . . . . .	87
<b>II</b>	<b>Activités de transfert et de valorisation</b>	<b>88</b>
6	Valorisation et transfert technologique . . . . .	89
6.1	Outils heuristiques pour la gestion des connaissances . . . . .	89
6.2	Logiciel d'analyse et de cartographie de corpus non structurés . . . . .	90
6.3	Gestion de communautés et des grandes organisations . . . . .	91
6.3.1	Community Explorer . . . . .	91
6.3.2	CNRS Explorer . . . . .	92
<b>III</b>	<b>Encadrement, animation et management de la recherche</b>	<b>94</b>
7	Responsabilités de direction de laboratoire . . . . .	95
7.1	Direction de l'Institut des Systèmes Complexes de Paris Île-de-France . . . . .	95
8	Responsabilités de coordination d'un réseau régional de recherche . . . . .	97
9	Responsabilités dans l'animation de programmes et projets de recherche . . . . .	99
10	Autres Responsabilités . . . . .	99

# **Première partie**

## **Parcours de recherche**

---

# 1 Régularités, lois et modèles en SHS

---

Mes recherches mobilisent régulièrement le concept de “modèle”. Avant d’en présenter les principaux axes et résultats, il me paraît important de préciser ce qu’il y a à modéliser en SHS, ce que recouvre le concept de modèle et à quoi cela sert. J’ai abordé ces questions dans quelques articles à visée épistémologique [Chavalarias (2008a,b,d,e); Deffuant et collab. (2015)].

## 1.1 Qu’y a-t-il à modéliser ?

Les décisions que nous pouvons prendre au niveau individuel ou collectif dépendent de nos anticipations sur l’évolution du monde qui nous entoure. Pour ne citer qu’un exemple, la sphère sociale et politique n’a pris au sérieux le problème du changement du climat – système dynamique complexe par excellence – qu’à partir du moment où un certain consensus s’est imposé sur les effets potentiels du réchauffement d’origine anthropique. La nature et l’étendue des sacrifices qui seront faits pour tenter d’infléchir ce dérèglement global seront à la mesure des effets escomptés.

Nous formons des anticipations au quotidien, souvent sans même nous en rendre compte. Celles-ci vont de simples anticipations corporelles – je tends la main car je m’attends à toucher cet objet –, à des raisonnements élaborés sur la relation entre l’état du monde et ses états futurs, entre nos actions et leurs conséquences, entre causes et effets. Pour penser l’articulation entre le présent et le futur, le schéma conceptuel du déterminisme laplacien a longtemps servi de guide. Cette notion héritière de la physique newtonienne relie l’instant présent de l’univers à son état futur de façon univoque et prédictible sous réserve d’une description parfaite de ses composantes. La conception du déterminisme qui lui est associée s’accommode des formes de hasard que nous rencontrons en les qualifiant de *hasard d’ignorance* : l’impression de hasard n’est que l’effet de notre ignorance concernant la multiplicité des causes. Dans ce cadre épistémologique, la connaissance parfaite du monde devient un horizon pour l’intelligence humaine qui, bien qu’elle ne puisse espérer l’atteindre, pourra s’en approcher asymptotiquement. Si l’on pouvait déterminer l’état du monde à un instant donné avec certitude, le hasard serait éliminé et l’homme connaîtrait son avenir.

Quel est le discours de la physique contemporaine sur cette notion de déterminisme ? S’il est indéniable qu’une meilleure description du monde nous permet de mieux le comprendre, la posture vis-à-vis de la question de la prédictibilité a changé de manière radicale.

### 1.1.1 Les enseignements de la physique

Le premier fait remarquable que l’on peut relever dans l’histoire récente de la physique est la distinction qui s’est opérée entre déterminisme et prédictibilité. Au siècle dernier, les travaux de mathématiciens tels que Henri Poincaré ont placé au centre des débats la question de la stabilité des systèmes dynamiques et le rôle de la mesure dans la prédiction. Une condition intuitive pour que l’on puisse prédire l’évolution d’un système dynamique en situation de connaissance imparfaite est que de petites erreurs sur la description du

système n'aient qu'une faible incidence qualitative sur notre prédiction. La question est donc de savoir si des systèmes aux descriptions proches ont des évolutions similaires. Or on montre que pour certains systèmes dynamiques, deux états qui seraient arbitrairement proches du point de vue de leur description donneraient lieu, à court ou moyen terme, à des évolutions radicalement différentes. On dit alors qu'il y a « sensibilité aux conditions initiales ». C'est le célèbre *effet papillon* de Lorenz [Lorenz (1972)], illustré souvent à tort par le battement d'ailes provoquant quelques semaines plus tard une tornade à l'autre bout du monde. Il est plus exacte de dire que le battement d'ailes de papillon est une cause partielle de la tornade (ou de son absence), qui devra tout autant sa manifestation au mouvement d'air que vous venez de provoquer en clignant des yeux. La grande découverte des années 1960 est que les systèmes dynamiques présentant des sensibilités aux conditions initiales, globalement ou localement (existence de singularités), ne sont pas pathologiques mais constituent au contraire la majorité des systèmes déterministes envisageables.

Cette découverte a entraîné un bouleversement épistémologique qu'il ne faut pas sous-estimer. Dans la perspective laplacienne, l'ignorance est toujours contingente. Mais le hasard d'ignorance est rendu quantitativement nécessaire du fait de l'instabilité de certains systèmes dynamiques : il y a toujours des causes qui, par leur petitesse, leur étendue ou leur multiplicité nous sont imperceptibles alors qu'elles engendrent à plus ou moins long terme des effets perceptibles. Ceux-ci nous apparaissent alors comme des événements imprévisibles dus au hasard. Par conséquent, la question de savoir si le monde est, ou n'est pas, déterministe devient indécidable : on sait que l'on ne pourra jamais savoir avec certitude.

Mais que dire alors de ces nombreux modèles qui dressent des portraits déterministes des phénomènes naturels ? Comme le montre Annick Lesne à partir de l'exemple du mouvement brownien Lesne (2008), il y a en fait une certaine indépendance entre la question de savoir si les phénomènes naturels sont déterministes ou stochastiques et celle de choisir une description du monde adaptée à nos besoins. En remontant les échelles d'observation du niveau moléculaire au niveau macroscopique, nous pouvons être amenés à penser un même phénomène tantôt de manière stochastique, tantôt de manière déterministe. La formulation de lois ou de régularités, qu'elles soient déterministes ou stochastiques, exige l'identification de causes et d'effets ou tout du moins, d'interdépendances de phénomènes, c'est-à-dire de structures évoluant de manière corrélée. Mais l'identification de structures dépend de l'échelle d'observation. Rien de plus désordonné qu'un gaz à l'échelle microscopique. Rien de plus ordonné qu'un gaz à l'échelle macroscopique, exemple parfait d'isotropie. Le problème n'est donc pas tant de savoir si un système est déterministe ou stochastique mais plutôt sur quelle fenêtre temporelle, avec quelle résolution spatiale et dans quels buts nous l'étudions.

Ce trait de la modélisation est une démarche générale que nous adoptons naturellement dans nos rapports quotidiens au monde. Le monde qui nous entoure est d'une infinie complexité du fait qu'il possède des structures à toutes les échelles. Nous sommes composés de cellules dont la quasi-totalité est renouvelée en l'espace de quelques semaines. Cela ne nous empêche pas de reconnaître quelqu'un dans la rue. La continuité d'un individu se situe pour nous à un autre niveau. Restreindre l'observation d'un phénomène à une certaine résolution et dans une certaine fenêtre temporelle permet de négliger les éléments qui interviennent à des échelles plus petites ou plus grandes et de penser le monde en termes de continuités et de relations causales. Au niveau d'observation considéré, nous savons que les imprécisions dans la mesure n'auront pas de conséquences mesurables, donc observables dans la fenêtre d'observation. Ainsi, bien que le principal axiome de la prédiction s'écroule, bien que des causes semblables ne produisent pas nécessairement

des effets semblables, bien qu'il n'y ait en fait jamais de causes semblables, nous pouvons faire chaque jour des expériences qui nous semblent relever d'une certaine causalité ou refléter un certain type de déterminisme.

Cette opération majeure de l'observateur qui, en négligeant certaines causes – mais nous avons vu qu'il n'a pas le choix –, définit un niveau d'observation, est à la fois ce qui nous permet d'identifier des continuités dans notre environnement et l'origine de cette impression occasionnelle de hasard radical. Ici encore, la physique et les formalismes sont porteurs de concepts riches pour caractériser ces niveaux d'observation. Nous avons évoqué le caractère imprédictible des trajectoires individuelles dans certains systèmes dynamiques dû à une incertitude sur les conditions initiales. Il y a cependant une contrepartie en termes de prédictibilité statistique avec le concept mathématique d'attracteur.

En isolant un niveau d'observation, il nous est parfois possible de simplifier par l'entendement les dynamiques à l'œuvre. Les entités que nous observons, si elles peuvent changer d'état, connaissent néanmoins des états plus ou moins stationnaires ou récurrents. Relativement à la précision spatio-temporelle fixée par l'observateur, les fluctuations des entités observées restent suffisamment petites autour d'une valeur moyenne pour être négligées statistiquement. L'ensemble des états ainsi considérés comme équivalents correspond à ce qu'on appelle formellement un attracteur (dans l'exemple ci-dessus, l'enveloppe corporelle est l'attracteur du métabolisme d'un individu ; c'est elle que nous reconnaissons). La propriété fondamentale d'un état appartenant à un attracteur est la suivante : un système évoluant à partir d'un état appartenant à un attracteur reste dans cet attracteur si les perturbations ne sont pas trop grandes.

En pratique, il est rare que l'on puisse prédire l'évolution d'un système dynamique avec exactitude. Néanmoins, il est généralement possible de déterminer qualitativement ses attracteurs. En déterminant les principaux attracteurs d'un système dynamique, il est ainsi possible de prévoir qualitativement l'ensemble de ses comportements futurs bien que l'on ne puisse prédire avec exactitude celui qui sera observé. C'est notamment ce que font les climatologues lorsqu'ils nous annoncent des scénarii possibles concernant l'évolution du climat pour les prochaines décennies. Nous ne pourrions jamais prévoir avec certitude ce qui risque d'arriver, mais nous pouvons essayer d'assigner des probabilités aux différents régimes climatiques à venir.

La réalité que nous percevons est ainsi constituée d'une multitude d'entités qui nous semblent avoir une certaine permanence précisément parce qu'elles ont atteint un attracteur relativement à notre *point de vue d'observation*. Les influences de leur environnement sont trop faibles à notre échelle pour les faire sortir de leur attracteur. Mais cette permanence, et donc la notion d'attracteur transposée à la réalité physique, est toujours relative à une échelle d'observation. Si nous nous coupons légèrement, notre corps cicatrise. Notre enveloppe corporelle, attracteur de notre métabolisme, n'est cependant qu'un état transitoire de la matière, que nous la regardions à l'échelle de l'espèce – où elle est éphémère –, ou à l'échelle de la cellule – où elle se renouvelle sans cesse. Les différents niveaux d'observation ont néanmoins une certaine légitimité dans la mesure où les grandeurs spatio-temporelles qui les caractérisent sont déterminées par les processus qui s'y déroulent : cycle de synthèse d'une protéine, cycle de vie d'une cellule, cycle de vie d'un organisme, cycles économiques, cycles climatiques, etc. La compréhension du couplage entre processus relevant de différents niveaux d'observation est une tâche difficile qui se trouve au cœur des approches « systèmes complexes ».

Les systèmes sociaux sont peut-être les systèmes pour lesquels ces enchevêtrements d'échelles d'espace et de temps sont les plus prononcés, et la tâche la plus ardue. Ici résident les sciences "très dures", selon le jeu de mot d'Herbert Simon. Pour nous atteler



à la compréhension et à la modélisation de ces systèmes, un détour par la biologie sera riche en enseignements.

### 1.1.2 Les enseignements de la biologie

Dans *Les Etincelles de Hasard* [Atlan (1999)], Henri Atlan reprend à son compte la citation de Szent-Györgyi placée en exergue de son ouvrage : « la vie n'existe pas ». Il peut sembler paradoxal que des biologistes en soient arrivés au point de nier l'existence de leur objet d'étude. Ce changement de perspective s'est opéré sous l'influence de la biologie moléculaire qui n'a cessé de démontrer que les briques élémentaires du vivant, auparavant jugées irréductibles à des propriétés physico-chimiques par les partisans du vitalisme, relèvent bien des lois de la matière inorganique. « Les mêmes lois s'appliquent, les propriétés seules varient : une pierre ne respire pas, une amibe ne pense pas... » [Atlan (2014)]. Mais par ce constat, Henri Atlan nous invite également à reconnaître la légitimité de la notion de vie comme catégorie d'expérience de notre vécu, en raison des propriétés particulières que les êtres vivants manifestent.

Le vitalisme est mort. La possibilité d'une réduction de l'organique à l'inorganique ne continue pas moins de poser des questions redoutables, malgré ce que laisserait penser le tout génétique de ces dernières années. Contrairement aux espérances soulevées par le séquençage du génome de plusieurs organismes, dont le génome humain, rien n'indique que « le livre de la vie » puisse se lire à l'aide des seules quatre lettres A, T, C, G<sup>1</sup>. Quelles en sont les raisons profondes ? Il semblerait que la boîte à outils conceptuels avec laquelle nous avons l'habitude d'approcher les objets inorganiques soit incomplète pour aborder le phénomène de vie. Si la vie n'est pas une notion explicative des propriétés organiques qui devrait être adjointe aux lois physiques, elle correspond néanmoins à un type d'organisation de la matière dont l'appréhension requiert une distinction entre réductionnisme « fort » et « faible ».

Considérons par exemple le problème de l'articulation du niveau génétique aux autres niveaux d'organisation. A contre courant de la majorité des travaux de ces dernières années, des études montrent aujourd'hui que la vie ne se réduit pas à l'exécution d'un programme inscrit sur une double hélice. Par exemple, dans le cas des eukaryotes (cellules à noyau), il y a des effets inter-niveaux dès la condensation de l'ADN en fibre de chromatine. Des études montrent que les interactions ADN-protéines sont radicalement différentes dans la chromatine (par rapport à l'ADN nu) du fait des contraintes mécaniques que la superstructure chromatinienne exerce sur l'ADN qui la constitue [Benecke (2006); Lesne et Victor (2006)]. Un autre exemple remarquable est la mise en évidence d'une relation entre les contraintes mécaniques s'exerçant sur les tissus cellulaires lors de leur croissance (pression physique sur les tissus) et l'expression des gènes dans les cellules qui les composent [Farge (2003)]. On met ainsi en évidence une rétroaction descendante entre un niveau physiologique et un niveau moléculaire. A chaque niveau d'organisation, les interactions entre les éléments de ce niveau sont donc susceptibles d'engendrer des superstructures qui vont avoir en retour un effet sur ces éléments en contraignant spatialement leurs interactions.

Nous voyons sur ces exemples que l'on ne peut comprendre le vivant sans prendre en compte l'enchevêtrement des niveaux d'organisation qu'il met en jeu. Un concept tout à

---

1. L'ADN de tout organisme vivant est "écrit" à partir de quatre molécules uniquement (bases nucléiques), symbolisées par les lettres A, C, G et T. Elles ont la capacité à s'apparier pour former la fameuse double hélice mise en évidence par Watson et collab. (1953)

fait essentiel pour appréhender ce phénomène et son lien avec les déterminismes physico-chimiques est le concept d'*émergence*, dont on peut trouver plusieurs acceptions dans la littérature. Un point commun à la plupart de ces définitions est l'idée d'apparition de structures macroscopiques sous l'effet d'interactions locales entre un grand nombre d'entités. Cette apparition peut ensuite être envisagée comme ontologiquement ou épistémiquement imprévisible, requérir ou pas un processus d'interprétation ou nécessiter un passage à la limite sur la taille du système considéré. Un cas d'école bien connu est celui de la fourmilière. Cet objet complexe se constitue à partir d'interactions locales entre des entités simples qui n'ont aucune sorte de représentation de leur comportement collectif.

En général, les niveaux d'organisation sont identifiables en tant que tels précisément parce que l'on y repère des structures émergentes. Que les phénomènes émergents aient une réalité ontologique ou soient liés à l'observation, une caractéristique importante du vivant est la présence de nombreuses rétroactions des structures émergentes sur les entités qui les ont engendrées, créant ainsi un bouclage entre différents niveaux d'organisation.

Quelques milliards d'années se sont écoulées depuis l'apparition des premières formes de vie. Depuis, la vie n'a cessé de tisser des liens complexes entre les différents niveaux émergents qu'elle articule. Des formes d'organisation, arrangements improbables entre entités hétérogènes, se sont stabilisées puis répliquées. Il n'y a là aucune sorte de finalité, mais des régularités qui s'auto-constituent à partir de processus émergents.

Rien ne vaut un exemple concret pour préciser le concept de *relations inter-niveaux* (causalités ascendantes et descendantes). Considérons le modèle de l'automate de tessellation [Bourgine et Stewart (2004); Stewart (2008)]. Inspiré des travaux de McMullin et Varela (1997), il conceptualise les propriétés auto-réparatrices d'une membrane délimitant un espace clos (une "cellule"). Nous pouvons le décrire de manière succincte de la façon suivante (cf. fig. 1) : un espace confiné (poche membranaire), plongé dans un substrat liquide composé de molécules  $A$ , est délimitée par une membrane composée de constituants  $C$ , qui se dégradent de manière aléatoire et doivent être remplacés régulièrement. Cette réparation ne s'effectue correctement qu'à condition que la concentration d'un substrat  $B$  soit suffisamment élevée dans la poche de liquide,  $B$  étant capable de se fixer sur la membrane pour la réparer. La surface interne de la membrane catalyse une réaction chimique  $A + A \rightarrow B$  conduisant à la formation de  $B$  qui reste captif de la poche délimitée par celle-ci (la membrane est imperméable à  $B$ ). La membrane en tant que macrostructure est donc une condition nécessaire au maintien d'une concentration élevée de l'élément  $B$ , qui répare la membrane et assure sa pérennité. Nous avons ici un exemple typique de *fait stylisé*, avec un modèle permettant d'appréhender des bouclages inter-niveaux constitutifs du phénomène de la vie : engendrement de macro-structures à partir de processus micros (intégration, *émergence*) et rétroaction de ces macro-structures sur les processus micros (régulation, *immersion*). Ce modèle nous offre également un exemple d'articulation entre plusieurs échelles de temps : la durée de vie de la membrane (temps longs, dynamique lente), la durée de vie de ses constituants et des éléments du substrat (temps courts, dynamique rapide). C'est un système dynamique dit *système dynamique lent-rapide*.

Ce modèle ne suffit certes pas à expliquer le phénomène de la vie. L'automate de tessellation n'est pas vivant. Mais cela nous donne des heuristiques pour mieux comprendre le phénomène de la vie, et notamment l'importance de l'enchevêtrement de niveaux d'organisation. Dans cet exemple, le modèle permet de questionner l'existence d'une zone de l'espace des paramètres où une structure membranaire est viable : si la réaction  $A + A \rightarrow B$  est trop lente par rapport au rythme de dégradation de la membrane, toute la structure s'effondre. Il souligne le type d'explication requise pour comprendre

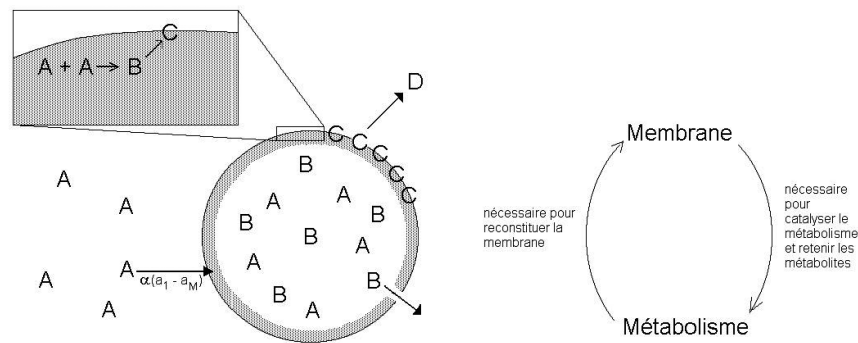


FIGURE 1 – Automate the Tesselation. A gauche : représentation de l'automate ; A droite bouclage inter-niveaux entre la membrane et le métabolisme. Source : Bourguin et Stewart (2004); Stewart (2008).

la stabilité de la membrane en tant que structure : il s'agit d'expliciter des processus et leur zone de viabilité, et non pas de simples relations d'interaction. Entre autres, nous comprenons que bien qu'il soit légitime de chercher à rendre compte des propriétés d'un *niveau d'organisation* en fonction des propriétés des entités du niveau inférieur (*réductionnisme faible*), il ne s'en suit pas qu'il soit légitime de chercher à expliquer l'ensemble des phénomènes d'après les propriétés d'un niveau unique (*réductionnisme fort*). On ne peut enchaîner les réductions successives comme on descendrait les marches d'un escalier. Cet exemple souligne la nécessité de conceptualiser les niveaux d'organisation dans la modélisation des systèmes complexes. Chercher à mettre "à plat" les relations entre leurs constituants, comme par exemple le proposent Latour et collab. (2012), relève de l'aporie conceptuelle.

Le lien subtil de dépendance/autonomie entre un niveau et le niveau supérieur à travers des processus de régulation illustré par l'automate de tessellation se reproduit à toutes les échelles du vivant. Des entités vivantes peuvent elles-mêmes entrer en interaction et prendre part à de nouveaux types de processus. C'est ainsi que les organismes vivants révèlent bien souvent une stratification en une multitude de processus enchevêtrés dont certains peuvent légitimement être qualifiés eux-mêmes de systèmes vivants. Les cellules de notre corps peuvent être cultivées en dehors de leur milieu d'origine comme des organismes vivants à part entière. Néanmoins, collectivement, par les processus qu'elles engendrent, elles définissent un nouvel organisme vivant qui se trouve être nous-même.

Cette organisation multi-niveaux est également omniprésente dans les systèmes sociaux : individu, cellule familiale, systèmes de parenté, groupes sociaux, organisations, institutions, etc. (voir par exemple Lazega (2016)). Comme nous l'avons déjà mentionné, les sciences humaines et sociales sont traversées de débats méthodologiques opposant holisme et individualisme méthodologique qui, si l'on y regarde de plus près, tourne précisément autour de la question des rétroactions ascendantes et descendantes entre niveaux d'organisation, chaque courant insistant sur la primauté d'un niveau sur l'autre. Par ce détour par la biologie, nous souhaitons suggérer une voix médiane qui envisagerait la co-constitution des différents niveaux, changeant radicalement le type de questions qu'il faut se poser. Nous verrons ainsi dans la section 4 qu'il est possible de développer bien plus qu'une simple analogie entre les concepts mobilisés par ces deux champs de sciences.

Le bouclage qui matérialise la double causalité ascendante et descendante que nous venons d'évoquer permet d'envisager des systèmes physiques dont l'activité stabilise la cohésion et la production de leurs propres composants. Que cette activité cesse et ces systèmes se désintègrent, ils « meurent ». Comme le fait remarquer Edgar Morin dans

*La Méthode*<sup>2</sup>, contrairement aux machines artificielles, exister et fonctionner sont deux modes inséparables des systèmes vivants. *L'identité de tels systèmes complexes n'est alors pas définissable par leurs constituants mais par les processus qui s'y déroulent et leur permettent de se produire eux-mêmes en permanence, leur caractère autopoïétique. Leur invariant fondamental est leur propre organisation.*

Cette manière de formuler le problème est le fil conducteur qui a guidé toute mon activité de modélisateur : chercher des processus plutôt que des relations, des attracteurs plutôt que des équilibres, articuler les niveaux d'organisation plutôt que les dissocier ou les opposer. Elle permet également de clarifier le statut des modèles et les contextes de leur utilisation.

## 1.2 La fonction du modèle

Venons-en maintenant à l'activité de modélisation proprement dite et à l'objectif que nous lui assignerons en sciences humaines et sociales. Commençons par évacuer ce qui pourrait être un contre-sens. Il ne s'agit pas ici de trouver de quelconques "lois" au sens où les physiciens peuvent chercher à découvrir les lois de la physique. Certains ont pu vouloir faire gagner leur discipline en scientificité en donnant le nom de "loi" à certaines approches formelles. Cette pratique est assez répandue en économie où l'on parle de "loi de l'offre et de la demande", "loi de l'utilité marginale", "loi de l'unicité du prix", etc., pratique qui a été vivement critiquée de l'intérieur même de la discipline [Boyer (2007); Guerrien (2004)]. Il est également fréquent que des modélisateurs venant de la physique donnent le nom de loi à des régularités qu'ils observent dans les données sociales. On parle par exemple souvent de "loi de puissance" ou "loi d'échelle".

D'après ce que nous venons de voir, il peut bien exister des lois en sciences humaines et sociales (notamment les lois de la physique), mais il faut les envisager avant tout comme des "conditions au bord", qui contraignent l'espace des possibles, mais laissent des possibilités infinies d'émergence de structures complexes et multi-niveaux, qui sont les *véritables objets sociaux*, et n'exhiberont des régularités que tant que les processus qui les maintiennent ne seront pas perturbés. L'étude de ces objets sociaux peut alors révéler les processus à œuvre dans leur émergence et les régularités qu'il engendrent, mais comme le soulignait déjà Stuart Mill, « la science de la société, considérée comme un système de déductions *a priori*, ne peut être une science de prédictions positives, mais seulement de tendances » (Mill (1843), liv. VI, chap. IX, § 2). Et l'observation de ces tendances devra toujours être assortie de l'estimation des échelles de temps sur lesquelles elles sont valides. Pour prendre une position extrême, à l'échelle de l'évolution des espèces, la plupart des régularités d'organisation sociale et de comportement observées sont éphémères et ne peuvent avoir le statut de loi. Mais sans aller jusque là, il est fréquent que des régularités qui ont pu être présentées comme des "lois" aient été détrônées de leur piédestal (cf. l'encart A pour un exemple).

Pour étudier les processus à l'œuvre dans la manifestation des phénomènes sociaux, plusieurs approches peuvent être mobilisées de manière complémentaire, en plus des méthodes traditionnelles de la sociologie.

1. **Sciences cognitives.** Comme le souligne Stuart Mill dans ce même ouvrage, "tous les phénomènes de la société sont des phénomènes de la nature humaine, produits

---

2. Morin et Morin (1977), page 193

par l'action des circonstances extérieures sur des masses d'êtres humains. Si donc les phénomènes de la pensée, du sentiment, de l'activité humaine, sont assujettis à des lois fixes, les phénomènes de la société doivent aussi être régis par des lois fixes, conséquences des précédentes." Cette démarche consistant à intégrer l'étude des caractéristiques individuelles dans l'analyse des classes de phénomènes sociaux a déjà été abondamment développée par de sociologues tels que [Tarde \(1890\)](#) - contemporain de Mill. Elle est pourtant restée marginale jusqu'à une période récente du fait de la compartimentation excessive des disciplines des sciences humaines et sociales et des rivalités quant à la définition de leurs frontières. Bien que l'on puisse penser que Durkheim "n'a cessé d'hésiter tout au long de son parcours sur la question des rapports individu/société, psychologie/sociologie, etc." [[Lahire \(2005\)](#)], les principaux courants de la sociologie sont longtemps restés méfiants vis-à-vis de la pertinence de la psychologie pour leur objet d'étude. De même, jusqu'à ce que Kahneman et Tversky aient le prix Nobel d'économie en 2002, l'économie "*mainstream*" est restée hermétique aux travaux issus de la psychologie, puis des sciences cognitives. Le développement rapide de ces dernières depuis les années 1990 est en train de changer la donne, et un nombre croissant d'économistes et sociologues admettent que l'étude des capacités cognitives individuelles doit faire partie intégrante de l'analyse des phénomènes socio-économiques.

2. **Analyse statistique, masses de données et l'apprentissage machine.** Puisque les processus à l'œuvre dans la constitution des faits sociaux se manifestent par la création de régularités et de tendances, les méthodes génériques visant à inférer des régularités ou des tendances directement à partir des données doivent faire partie de la boîte à outils du chercheur. Ceci explique l'incursion spectaculaire de l'apprentissage machine (*machine learning*), l'*apprentissage de représentation* et de la mode de du *deep learning* [[Bengio et collab. \(2013\)](#)] dans l'étude de certains phénomènes sociaux depuis l'avènement des *méga-données* en SHS (traces digitales des activités en ligne, données sociales numérisées rétrospectivement, etc.). Trop volumineuses pour pouvoir être traitées par les seules méthodes classiques d'investigation, l'application de ces méthodes à ces méga-données constituent de véritables opportunités pour établir les caractéristiques des processus sociaux ayant générés ces dernières. Il ne faut cependant pas se fourvoyer sur la permanence des tendances ainsi observées et une approche complémentaire par modélisation est nécessaire pour définir et estimer le domaine de validité de ces observations. Le coq fait la fête au fermier qui tous les matins lui donne à manger, ... jusqu'au jour du coq au vin.
3. **Les modèles et la modélisation.** Comme nous l'avons vu, les modèles sont essentiels dans la démarche d'explicitation des relations entre les différents niveaux d'organisation d'un système, et l'analyse des processus qui les maintiennent. Le concept de modèle recouvrant plusieurs types de pratiques, il convient de les détailler.

Nous considérerons le terme "modèle" dans son sens associé à l'activité de modélisation<sup>3</sup> : "*une représentation d'un dispositif réel par un système formel - par exemple un*

3. "On retrouve dans les définitions des dictionnaires contemporains, 5 classes de signification et d'usage : le modèle comme référent ou prototype à reproduire (le modèle du peintre), le modèle comme maquette d'un dispositif réel (chez le fondeur, l'architecte), le modèle comme type idéal dégagé d'une population homogène (un modèle de sainteté, de candeur, de beauté...), le modèle comme icône ou dispositif mécanique représentant une idée abstraite (le modèle hydraulique de la circulation monétaire chez Irving

*système d'équations - qui permette à la fois de penser ce dispositif, d'étudier son fonctionnement, et d'agir sur lui pour le contrôler ou le faire évoluer*" [Armatte (2005)]. Cette définition de "modèle" est héritée du courant positiviste du XIX<sup>ème</sup> siècle qui, concevait la démarche de modélisation comme la définition d'un isomorphisme entre un système naturel et un système formel. L'horizon de cette démarche se devait d'être la prévision parfaite du comportement du système naturel, au sens de Laplace. Mais la notion contemporaine de modèle s'est enrichie d'autres acceptions, notamment au sein des sciences humaines et sociales, en même temps que les pratiques de modélisation se diversifiaient.

Devant la complexité des systèmes socio-économiques, la fonction du modèle a été étendue de la description exacte de phénomènes réels à celle de "*faits stylisés*", c'est à dire de tendances ou observations empiriques robustes au niveau macro, non nécessairement quantifiées de manière précise. Cette approche est très populaire en économie - pour ne citer que quelques faits stylisés : "*l'usage dans le temps des nouvelles technologies suit une courbe en S*", "*il y a un lien positif entre consommation d'énergie et croissance économique*", "*les cycles économiques sont périodiques*", etc.. - ; et en éconophysique et sociophysique, avec par exemple de nombreuses *lois de puissance* découvertes puis reproduites dans le cadre de modélisations. L'activité de modélisation consiste alors à proposer des modèles qui approximent certaines caractéristiques d'un système naturel en reproduisant certains faits stylisés. Lorsqu'elle est réussie, cette démarche permet de formuler des hypothèses sur les mécanismes à l'origine de la manifestation de ces faits stylisés dans le système étudié. Si l'existence de ces mécanismes est confirmée par l'observation du système, le modèle peut alors prétendre à une explication des faits stylisés et formuler des prédictions sur les conséquences possibles d'intervention sur ces mécanismes.

Un exemple paradigmatique de cette démarche de modélisation en sciences humaines et sociales est le modèle de ségrégation de Schelling (1978). À partir du constat de la persistance d'une ségrégation urbaine entre des différentes communautés ethniques ou classes sociales (fait stylisé que l'on retrouve dans beaucoup de villes) plusieurs hypothèses peuvent être émises : cette ségrégation est organisée *top-down* par les pouvoirs publics, elle est le résultat de la composition *bottom-up* d'attitudes discriminatoires (raciales ou économiques), elle est l'effet des différences de langues, etc.

Le modèle de Schelling quant à lui démontre l'effet contre-intuitif des seuils de tolérance individuels dans le choix d'un lieu d'habitation. Le mécanisme dont il a fait l'hypothèse, et dont l'importance peut être mesurée empiriquement, est qu'en présence de deux groupes sociaux distincts *A* et *B*, chaque personne possède un seuil de tolérance quant à la proportion minimale de voisins du même groupe qu'elle (par exemple chacun veut pouvoir parler sa langue natale avec au moins quelques voisins). Chaque membre de ces groupes est donc prêt à déménager pour un logement lui permettant de satisfaire cette préférence si la proportion des voisins du même groupe que lui passe en dessous d'un certain seuil.

Schelling a démontré que ce mécanisme suffit à générer de la ségrégation de population en quartiers, quand bien même ce seuil serait bas (par exemple 30% dans l'exemple de la figure 2). Cet effet macro d'un mécanisme micro est hautement contre-intuitif et ne peut être étudié sans modélisation. Cette modélisation n'a pas pour prétention de décrire avec exactitude les phénomènes de ségrégation, ni même d'affirmer que ces mécanismes sont effectivement à l'œuvre, ce qui relève d'un travail de terrain. Elle ouvre en revanche de nouvelles directions de recherche pour l'explication de ce phénomène dans de nombreuses situations où un tel mécanisme peut être identifié (écoles, quartiers, loisirs, etc.).

---

Fisher par exemple), le formalisme logico-mathématique qui représente un système." Armatte (2005)

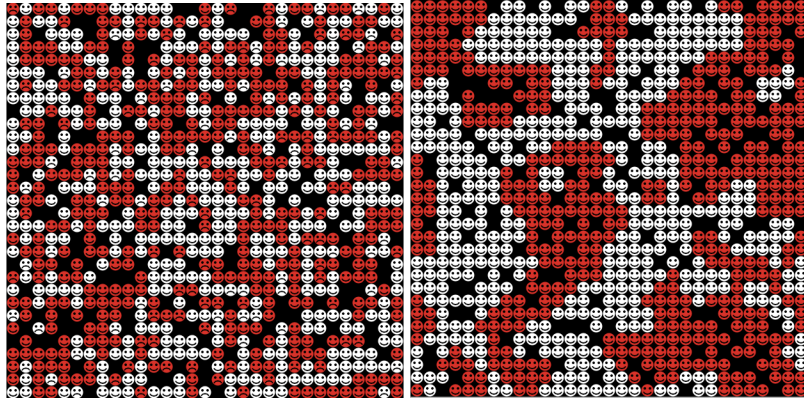


FIGURE 2 – **Exemple de fait stylisé : phénomènes de ségrégation urbaine.** Test de l’hypothèse de Schelling sur une société artificielle. Des agents sont placés aux nœuds d’une grille torique. Chaque agent est satisfait si et seulement si au moins 30% de voisins sont de la même couleur que lui (parmi les 8 cellules adjacentes). Dans le cas contraire, il se déplace aléatoirement sur une case vide (en noir). Le processus initié avec une distribution aléatoire (à gauche) converge très rapidement vers une distribution avec ségrégation (à droite). Simulation réalisée sous NetLogo [Wilensky (1997, 1999)]

Le modèle a, dans ce cas précis, une fonction éminemment heuristique, la fonction descriptive étant placée au second plan, ce que revendiquait Schelling (1978) :

Simplified models of artificial situations can be offered for either of two purposes. One is ambitious : these are “basic models”—first approximations that can be elaborated to simulate with higher fidelity the real situations we want to examine. The second is modest : whether or not these models constitute a “starting set” on which better approximations can be built, they illustrate the kind of analysis that is needed, some of the phenomena to be anticipated, and some of the questions worth asking. The second, more modest, accomplishment is my only aim in the preceding demonstrations.

Ce nouveau rôle du modèle, qui a pu être qualifié de “médiateur” entre la théorie et les données [Morgan et Morrison (1999)], est éminemment structurant dans des domaines scientifiques pour lesquels la difficulté de quantifier rend parfois les différentes théories incommensurables. Loin de mutiler les approches qualitatives et de contraindre leur point de vue, les modèles permettent au contraire de mieux en exprimer les spécificités à des fins de débat scientifique, ce que souligne par exemple Jean Ullmo dans la préface de l’un de ses ouvrages [Ullmo (1969)] :

Les modèles sont indispensables, et il faut les multiplier. Chacun dans son formalisme mathématique rigoureux, permet de délimiter des hypothèses, de préciser des concepts, d’expliciter des conséquences, de reconnaître des convergences ou des incompatibilités. Ils abolissent donc les environnements indécis, les glissements de pensée, les actions contraires aux fins proclamées, les incohérences de projet ou d’acte. Leur multiplicité s’oppose à la tentation du mécanisme unique, de la solution parfaite : elle rappelle que l’extraordinaire complexité des actions et réactions économiques ne peut être passible d’un modèle exhaustif [...] Les modèles sont donc des instruments de pensée. On ne peut s’en passer, on ne doit pas s’y soumettre.

Au-delà de l’élaboration d’une connaissance scientifique, les modèles et l’activité de modélisation sont également utilisés dans un cadre pragmatique et décisionnel comme

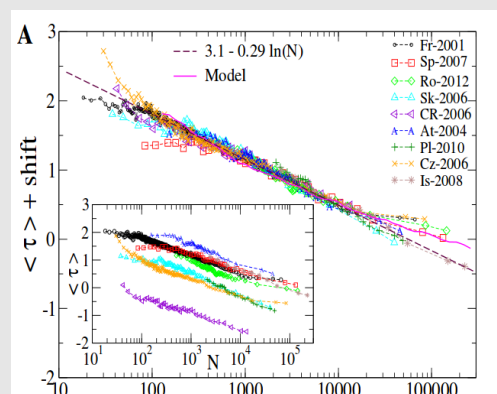
supports à la confrontation des points de vue au cours de processus de prise de décisions collectives. C'est ce que développe par exemple l'approche dite de "modélisation d'accompagnement" (COMMOD) [Barreteau et collab. (2003)], qui conçoit l'activité de modélisation comme une confrontation itérative et permanente entre les théories et les observations de terrain, à l'aide de simulations informatiques ou même de jeux de plateau. Dans ce cadre, l'activité de modélisation, par les échanges qu'elle induit entre les différents protagonistes d'un problème, est au moins aussi importante que le modèle lui-même et sa validité.

L'évolution de la notion de modèle et des pratiques de modélisation en sciences humaines et sociales reflète une évolution profonde du rapport à la connaissance et de l'épistémologie développés par toute une communauté interdisciplinaire de recherche, qui s'est en partie cristallisée autour de l'étude des systèmes complexes.



## A. Loi d'échelle dans les comportements électoraux : lorsque la France brise l'universalité

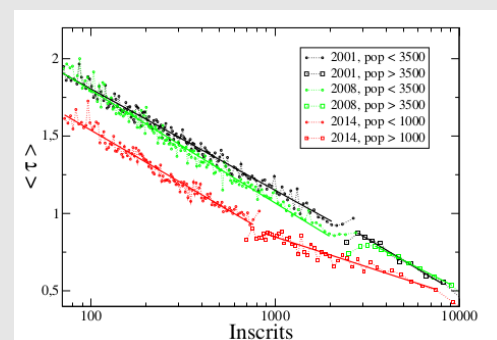
Les élections sont depuis les années 1990 un terrain privilégié pour la modélisation de comportements sociaux. Ces processus de décision collective s'opèrent dans des conditions bien définies (la réglementation électorale), donnent lieu à la collecte massive de données rendues publiques (le résultat des votes) et ont pu être observés de manière répétée dans de nombreux contextes. Un des faits stylisés très connu pour les élections locales est le fait que le taux d'abstention croisse avec le nombre d'inscrits sur les listes électorales (et donc la taille de la ville), les grandes villes connaissant une plus grande abstention que les petites villes pour une même élection. Pour essayer de quantifier et modéliser cet effet, [Borghesi et collab. \(2013\)](#), ont analysé les résultats d'une vingtaine d'élections municipales dans dix pays différents avec différentes caractéristiques sociales, économiques et culturelles. Ils ont ainsi montré qu'il y avait un "effet universel de taille" (*universal size effects*) caractérisé par le fait que le taux logarithmique moyen de participation, défini par  $\tau = \langle \ln \frac{N_+}{N_-} \rangle$  (où  $N = N_+ + N_-$  est le nombre total d'inscrits avec  $N_+$  nombre de votants et  $N_-$  nombre d'abstentions), suit une fonction linéaire  $\tau \approx C - \alpha \ln(N)$ ,  $\alpha$  étant un paramètre "universel" (cf. figure ci-dessous).



Ce type de régularité est très important pour interpréter les résultats de l'abstention ville par

ville (systématiquement commentés dans les médias), car cela suggère qu'il faille normaliser les taux d'abstention en fonction de la taille de la ville avant d'interpréter les résultats. C'est ce que nous avons fait dans un projet de plate-forme web qui permet d'explorer les résultats des 1er tours des élections municipales 2001, 2008 et 2014 en France, et sensibiliser le public à cette problématique [Chavalarias et collab. \(2014\)](#).

Mais en analysant avec la même méthodologie la participation au 1er tour des municipales 2014, nous avons observé que cette régularité a été brisée (cf. figure ci-dessous) pour les communes de moins de 1000 habitants environ. La droite de 2014 est parallèle aux droites des élections précédentes mais décalée vers le bas, traduisant une diminution globale de la participation. En revanche, pour les petites communes ( $N \leq 1000$ ) cette diminution de la participation avec la taille de la commune est nettement moins prononcée. Il faut noter qu'en 2014 les communes de plus de 1000 habitants sont passées à la règle électorale générale. Cependant les régularités observées précédemment ne dépendaient pas de la règle électorale.



Une des interprétations possibles de cette brisure de régularité est une modification profonde du paysage politique français en terme de processus de recrutement des électeurs ou d'équilibre des partis. Le fait que les petites municipalités aient voté au-delà de ce qui était attendu des régularités précédentes pouvait être le signe d'une modification des comportements électoraux en zone rurale, ce qui a d'ailleurs été observé aux régionales 2015 avec la montée spectaculaire du FN, qui y a recueilli 41% des voix. Le concours de disciplines spécialisées (sociologie, histoire, sciences politiques) est essentiel pour poursuivre cette analyse au delà de la simple spéculation. Nous avons néanmoins ici un exemple où un phénomène social semble obéir à une loi ... jusqu'à ce que celle-ci change.

### 1.3 L'étude des modèles

La formulation d'un modèle ne suffit pas à répondre à une question. Elle pose même parfois plus de problèmes qu'elle n'en résout car il faut bien étudier les comportements du modèle pour obtenir des informations sur le système qu'il modélise. Jusqu'à la fin du XX<sup>ème</sup> siècle, la résolution analytique d'équations mathématiques a été la règle d'or de cette exploration des modèles, bien souvent au prix de l'adaptation du modèle à ce que sont en mesure de traiter les mathématiques (hypothèse de l'agent représentatif, du champ moyen, de panmixie, de l'espace infini, de l'état asymptotique au temps infini, de l'état d'équilibre, etc.).

A mesure que les effets de ces hypothèses simplificatrices se sont révélées en décalage avec la réalité à étudier, qu'ont été introduits le caractère hétérogène des agents, la finitude des populations, la topologie des interactions (jeux spatiaux, interactions en réseaux ou dans des espaces n-dimensionnels continus), etc., de nouvelles méthodologies d'exploration des modèles sont apparues autour de la simulation informatique, des automates cellulaires et des systèmes multi-agents. Celles-ci se sont développées d'autant plus vigoureusement qu'elles ont été favorisées par l'accroissement exponentiel des capacités de calcul (la plupart des ordinateurs de bureau actuels sont plus puissants que les super-calculateurs que la NASA a utilisé pour envoyer les premiers hommes sur la lune). Nous avons déjà vu section 1.1.1 que diverses méthodologies peuvent coexister pour l'étude d'un même phénomène en fonction du niveau d'observation adopté. Même si la communauté des mathématiciens a pu se montrer sceptique devant ces méthodes qui ne proposent pas de théorème, ou que des partisans de la simulation informatique ont pu proclamer l'avènement d'un "nouveau type de science" qui pallierait aux échecs des approches mathématiques [Wolfram (2002)], la complémentarité de ces deux approches reste réelle.

Le recours à la simulation informatique est néanmoins incontournable lorsque l'on cherche à étudier certains phénomènes qui mettent en scène des interactions spatialisées. Cela pour des raisons théoriques. Dans la continuité des travaux de Gödel démontrant l'incomplétude des langages formels, et de ceux de Church et Turing démontrant l'insolvabilité ou l'indécidabilité de certains problèmes formels Church (1936a,b); Turing (1936), il a été démontré que certains systèmes physiques sont *computationnellement irréductibles* Wolfram (1985) :

[...] many physical systems are computationally irreducible, so that their own evolution is effectively the most efficient procedure for determining their future. As a consequence, many questions about these systems can be answered only by very lengthy or potentially infinite computations. But some question answerable by simple computation may still be formulated.

Pour l'étude de ces systèmes, il n'y a pas de plus court chemin que d'observer leur évolution ou d'utiliser un modèle pour reproduire *in silico* une évolution comparable. Ce qui est vrai pour les systèmes physiques l'est *a fortiori* pour les systèmes plus complexes tels que les systèmes biologiques ou les systèmes sociaux, dont les principes d'évolution incluent les lois de la physique.

Prenons l'exemple de la théorie des jeux évolutionnaires, qui sera le cadre de plusieurs de nos approches de modélisation. Dans les modèles adoptant cette perspective, des po-

pulations d'agents évoluent suivant des processus de sélection qui tiennent compte des interactions entre tous les agents. Les agents possèdent différents types et stratégies qui leur permettent de gagner ou perdre des points de *fitness* lors de leurs confrontations avec les autres agents. Au cours du temps, ces stratégies prolifèrent ou perdent du terrain, l'un des enjeux de l'approche par modélisation étant de prédire quelle(s) stratégie(s) prendra le dessus (quelles espèces vont envahir un écosystème ? dans quels environnements des attitudes pro-sociales sont elles viables ? etc.).

La grande majorité des modèles qui exhibent une résolution analytique font l'hypothèse de la panmixie : les individus sont répartis de manière homogène au sein de la population et se reproduisent au gré d'appariements aléatoires. Cependant, les résultats des modèles sont totalement différents si l'hypothèse de panmixie est remplacée par des interactions spatialisées [Nowak et Sigmund (2000)]. Bien que dans quelques cas simples, il soit possible d'obtenir une résolution analytique de ces modèles [Dieckmann et collab. (2000)], ceux-ci sont le plus souvent étudiés à l'aide de simulations informatiques et non sans raison : on peut construire des exemples (par exemple autour du dilemme du prisonnier spatialisé) tels que la question de savoir quelles stratégies sont viables ou invasives est indécidable [Grim (1994)]. De même que la physique a découvert que les systèmes exhibant une sensibilité aux conditions initiales étaient la règle plutôt que l'exception, il ne serait pas surprenant, comme le soutient d'ailleurs Wolfram (2002), que la question de prédire l'évolution d'un système dynamique spatial soit en général indécidable, ou du moins NP-complet<sup>4</sup>.

Nous pouvons alors mesurer l'écart entre cette conception du modèle et sa conception traditionnelle :

“le modèle n'est plus un système formel qui représente un système physique bien défini et délimité pour lequel on a des lois valables *ceteris paribus*. Le modèle devient un mécanisme d'intégration de données produites par des sous systèmes d'information. Le modèle, ensemble d'équations, d'interfaces, de données qui s'est transformé en logiciel, constitue un système de substitution au système réel, dont on n'a pas de théorie complète, et qui permet de faire des expériences fictives, pour comprendre le jeu complexe de ses interactions. Et ces expériences fictives constituent bien une méthodologie de rechange, par rapport à la méthode hypothético-déductive aussi bien que par rapport à la méthode expérimentale, dans le cas des systèmes complexes.”  
Armatte (2005)

## 1.4 Reconstruction théorique, reconstruction phénoménologique

Pour finir de camper le décor de la modélisation, dans lequel évoluent mes travaux, je souhaiterais revenir sur un schéma générique pour décrire la démarche de modélisation, introduit dans le chapitre “*Épistémologie formelle, expérimentation, apprentissage automatique*” (Bourgine P., Brodu N., Deffuant G., Kapoula Z., Müller J-P., Peyreiras N.) de la feuille de route des systèmes complexes [Chavalarias et collab. (2009)].

Il y est fait la distinction entre *reconstruction théorique* et *reconstruction phénoménologique*.

4. Les problèmes NP-complets sont les problèmes pour lesquels il n'existe pas de résolution en temps polynomial, c-a-d que leur résolution prend un temps qui s'allonge exponentiellement avec la taille des données, dépassant rapidement toute capacité de calcul possible.

### 1.4.1 Reconstruction théorique

La *reconstruction théorique* consiste à proposer des modèles (systèmes formels ou computationnels) qui synthétisent la diversité phénoménale et ont des propriétés génératives qui leur permettent d'expliquer des phénomènes passés, de prédire des faits futurs ou de reproduire des structures observées. C'est le cas par exemple des *L-systèmes*, grammaires formelles inventées en 1968 par le biologiste hongrois Aristid Lindenmayer, dont le but est de modéliser le processus de développement de plantes ou de prolifération de bactéries (cf. fig. 3). Une fois que l'on a inféré les paramètres d'un *L-système* qui reproduit le plus fidèlement une espèce de plante donnée, il est possible de simuler la croissance des individus de cette espèce et les hétérogénéités inter-individuelles en faisant varier quelques paramètres seulement. Les *L-systèmes* sont très utilisés par exemple dans les films d'animation pour simuler des paysages réalistes.

Un autre exemple de reconstruction théorique peut-être trouvé dans les pratiques de modélisation du changement climatique qui intègrent des connaissances très diverses à propos d'un grand nombre de sous-systèmes dans le but de faire des prédictions d'évolution du climat sous différents scénarii.

Le lecteur aura reconnu sous le concept de reconstruction théorique les concepts familiers de modélisation formelle et computationnelle.

### 1.4.2 Reconstruction phénoménologique

La *reconstruction phénoménologique* est une démarche consistant à pré-structurer les données en vue d'une compréhension plus globale lors de traitements ultérieurs. Elle est rarement pensée comme un pré-requis à la reconstruction théorique. De fait, le nom de "données", sur lesquelles les modèles sont validés, suggère des entités directement accessibles à l'expérience et qui n'ont pas besoin de traitement particulier pour être intégrées à une entreprise conceptuelle. Il suffit de mesurer.

La reconstruction phénoménologique est pourtant plus courante qu'on ne le pense. Il suffit, littéralement, d'ouvrir les yeux. L'illusion d'unité que nous ressentons lors de la perception de scènes visuelles, avec leurs objets aux contours bien définis, est en fait une reconstruction qui commence dès la réception du signal lumineux par la rétine. Le système visuel détecte en effet les contours des objets grâce à une multitude d'extrapolations très bas niveau d'unités élémentaires de contour, calculées à partir des signaux rétiniens. Il le fait en respectant certaines règles de continuité et de lissité déjà décrites au

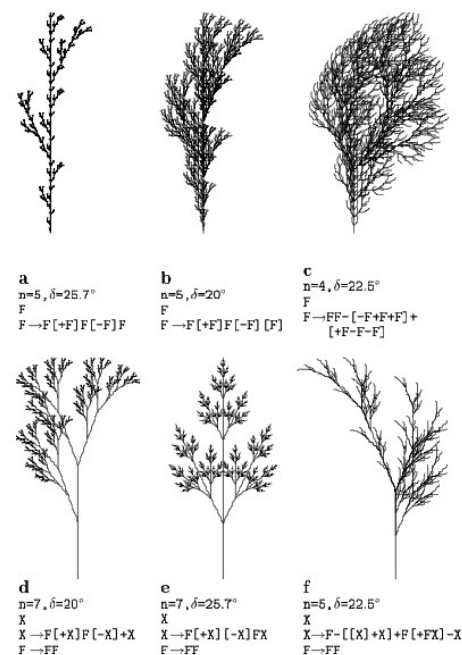


Figure 1.24: Examples of plant-like structures generated by bracketed OL-systems. L-systems (a), (b) and (c) are edge-rewriting, while (d), (e) and (f) are node-rewriting.

FIGURE 3 — Exemple de reconstruction théorique. Exemple de L-système d'Aristid Lindenmayers simulant la croissance de plantes.

début du XXème siècle, comme d'autres phénomènes de perception des contours, par la loi gestaltiste dite de "bonne continuation".

Il a été démontré [Bressloff et collab. (2002); Petitot et Tondut (1999)] que cette capacité de détection des contours est une propriété de l'architecture neuronale des aires primaires visuelles (architecture en pinwheels des hyper-colonnes de V1), c'est-à-dire de la manière dont sont organisés spatialement et reliés entre eux les neurones des aires du cerveau dédiées à la vision. C'est une capacité qui tient du "hardware" et peut être modélisée à l'aide de structures mathématiques (fibré de contact, Petitot (2003)).

Nous pouvons faire l'expérience de cette reconstruction en regardant la figure 4. La grande majorité des gens y voient deux triangles superposés, l'un ayant une teinte de blanc légèrement différente du fond. On peut "voir" les contours du triangle supérieur et même les suivre du doigt. Et pourtant, par construction, ils n'existent pas, car la figure ne contient que les trois angles et les trois camemberts avec deux uniques couleurs : noir et blanc. Les contours illusoire sont une pure reconstruction de notre système visuel qui est "câblé" pour nous permettre d'interpréter au mieux les scènes visuelles, notamment lorsque certaines parties d'objets sont occultées.

Ces recherches sont très importantes pour comprendre la *perception*, premier maillon de la chaîne de l'interprétation et de la constitution du sens. Elle est le couplage entre un stimulus et un processus de filtrage/pré-structuration conditionné par notre expérience passée, dont le résultat sert de prémices à des interprétations plus haut niveau. Il a par exemple été découvert que le type de contours auquel le système visuel d'un organisme est sensible dépend des types d'environnements auxquels son cerveau a été exposé dans les premières années de son développement [Sengpiel et collab. (1999)]. L'architecture neuronale fonctionnelle, qui résulte de cet apprentissage, contraint les dynamiques neuronales qui mènent à la perception des formes. Elle contraint la perception d'un individu tout au long de sa vie. Cette architecture peut être mise en évidence via ses *modes propres*, qui expliquent certaines hallucinations visuelles et peuvent être "prédits" à partir d'un modèle du cortex visuel [Bressloff et collab. (2002)].

Cette opération de pré-structuration des données, dont l'activité du cortex visuel nous donne un exemple, est appelée *reconstruction phénoménologique*. Lorsque l'on s'intéresse à la modélisation des systèmes complexes, elle devient une étape fondamentale de l'analyse en raison du caractère foisonnant et non structuré des données qu'ils génèrent. Pensez par exemple à la tâche consistant à délimiter de manière *bottom-up* les champs scientifiques à partir des millions d'articles produits annuellement. C'est une étape préliminaire si l'on se pose des questions sur l'appartenance d'un ensemble de travaux à un champ, la question des régularités dans l'évolution des sciences, l'influence des champs de recherche sur les activités des chercheurs, etc. Pourtant cette tâche est hautement non triviale. Elle constitue l'un de mes axes de recherche, présenté en section 3.2. Nous pouvons d'ores et déjà annoncer qu'il n'y a pas qu'une seule manière de délimiter des champs

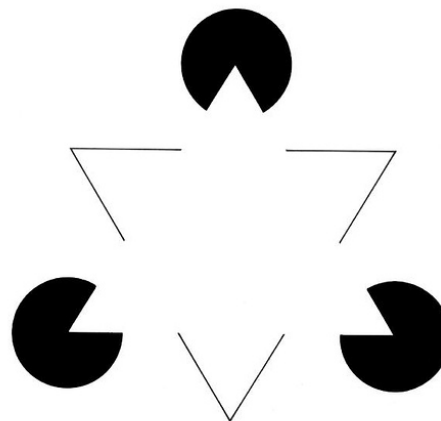


FIGURE 4 – Exemple de reconstruction phénoménologique. Illusion dite des triangles de Kanizsa Kanizsa (1979), conséquence de l'architecture du cortex visuel, "câblé" pour détecter les contours.

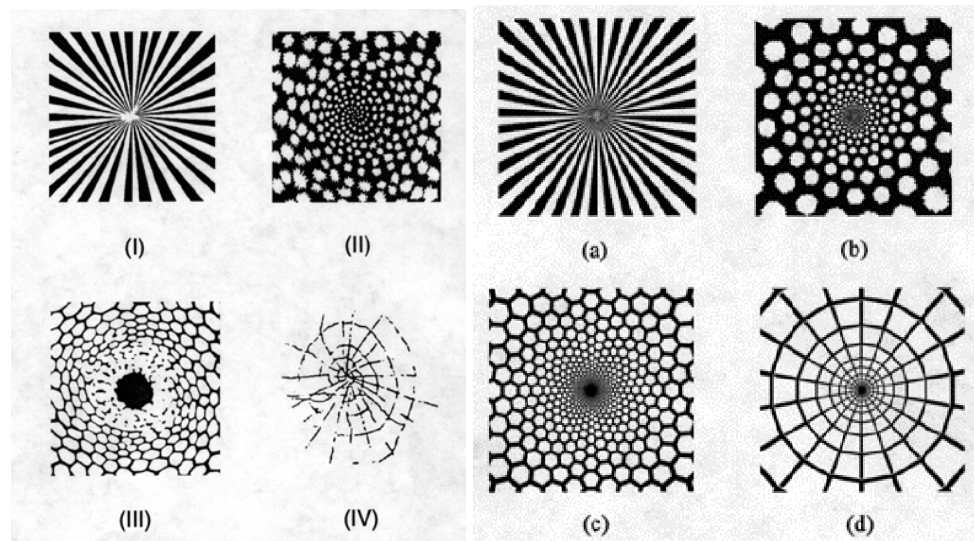


FIGURE 5 — Hallucinations reconstruites grâce à la modélisation de l'architecture du cortex visuel. A gauche, classification par Heinrich Klüver (1926) d'hallucinations dessinées par des patients sous psychotrope. A droite, modes propres prédits par un modèle de l'architecture des aires visuelles primaires [Bressloff et collab. (2002)].

scientifiques. Définir des algorithmes pour automatiser cette tâche implique de poser un certain nombre d'hypothèses et de proposer un modèle de ce qu'est un champ. Ces hypothèses engendreront nécessairement des "contours illusoires" comme dans les triangles de Kanizsa. Mais la fonction de ces contours ne sera pas tant de représenter des contours objectifs et existant dans le monde réel que de nous permettre de pré-structurer les données pour pouvoir ensuite concevoir des modèles du monde qui font sens.

### 1.4.3 Articulation des opérations de reconstruction

Les concepts regroupés sous les appellations de *reconstruction théorique* et *reconstruction phénoménologique* ne sont pas nouveaux, ils recouvrent les activités de modélisation formelle ou computationnelle et d'inférence de motifs (aussi appelée fouille de données). Il est néanmoins opportun de les rassembler sous l'appellation de *reconstruction* pour souligner leur articulation suivant les trois étapes de la modélisation :

1. identifier les entités pertinentes pour une échelle spatio-temporelle donnée,
2. caractériser les interactions entre les entités et les processus qu'elles engendrent,
3. évaluer et formaliser le comportement du système.

1) relève de la reconstruction phénoménologique, 3) de la reconstruction théorique et 2) des deux.

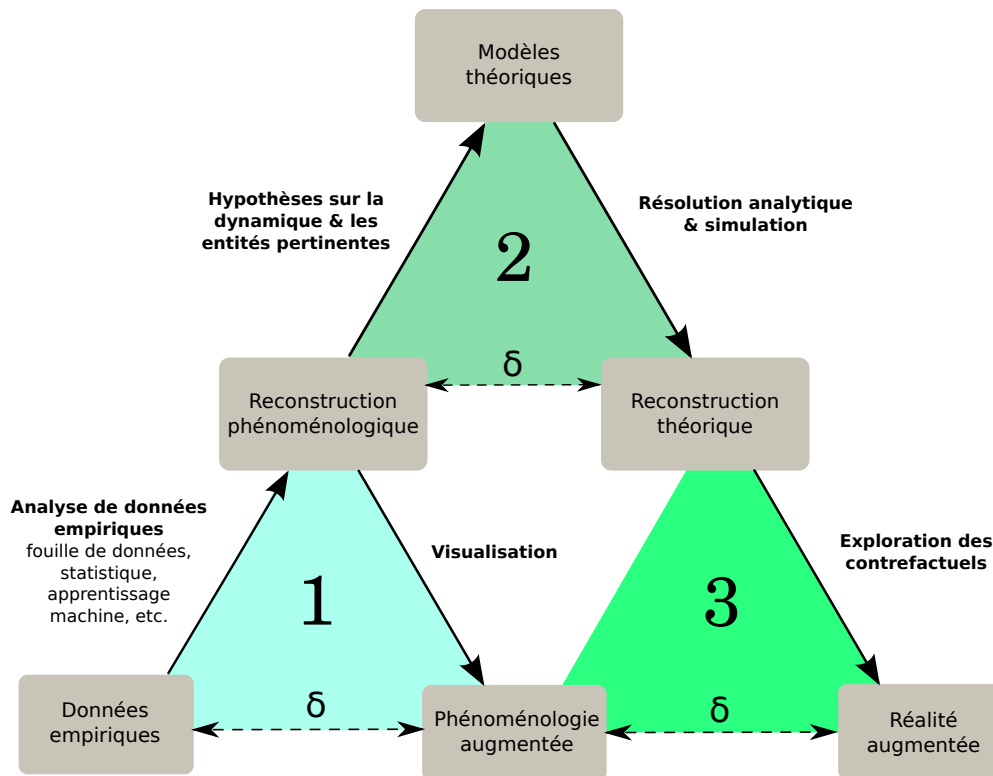


FIGURE 6 – Le schéma ci-dessus [source Chavalarias et collab. (2009)] représente l'articulation des opérations de modélisation suivant trois domaines d'activité : reconstruction phénoménologique, reconstruction théorique et élaboration de dispositifs d'interaction avec les données produites par ces deux dernières. Ces opérations se complètent mais peuvent chacune constituer des activités de recherche à part entière et faire l'objet d'études indépendantes (on peut faire de la fouille de données sans modèle, de la modélisation sans données ou des outils de visualisation génériques). Une démarche de modélisation exhaustive articulera cependant ces trois composantes en faisant intervenir l'humain dans la boucle pour évaluer les écarts ( $\delta$ ) des différentes reconstructions avec ce qui fait sens pour lui dans le monde. Ces écarts peuvent aussi être évalués via des méthodes mathématiques ou computationnelles à chaque fois qu'il s'agit de mesurer des différences entre structures (par exemple entre un réseau dynamique issu d'une reconstruction phénoménologique et sa reconstruction théorique.)

## 2 Principaux objets de recherche

*Les premières années de mon activité de chercheur (jusqu'à mon premier post-doc), ont été essentiellement consacrées à la reconstruction théorique de dynamiques culturelles (triangle 2 de la figure 6). Mon activité de recherche s'est par la suite étendue aux triangles 1 (reconstruction phénoménologique & visualisation) puis 3 (exploration interactive de données et question de la prédiction), l'objectif étant de mettre en œuvre ces trois opérations à propos d'un même objet.*

Le terme qui qualifie le plus justement mon parcours intellectuel est dans doute “*interdisciplinarité*”. Cet intérêt pour la recherche aux interfaces entre les disciplines se reflète dans ma formation, mes collaborations et la diversité disciplinaire des revues et ouvrages dans lesquels j'ai eu l'occasion de publier. Il est intrinsèquement lié à des objets d'étude qui n'ont cessé de me fasciner tout au long de mon parcours intellectuel : la complexité des sociétés humaines et la diversité des cultures qu'elles développent.

Cette omniprésente complexité, grâce à laquelle *homo-sapiens* a traversé les âges conquérant toujours plus territoires au détriment des autres espèces, est comme le faisait justement remarquer Schelling (1978), l'un des miracles de la Nature :

We expect this fantastically complex system [our society] to be even better coordinated than it sometimes is. Tens of millions of people making billions of decisions every week about what to buy and what to sell and where to work and how much to save and how much to borrow and what orders to fill and what stocks to accumulate and where to move and what schools to go to and what jobs to take and where to build the supermarkets and movie theatres and electric power stations, when to invest in buildings above ground and mine shafts underground and fleets of trucks and ships and aircraft - if you are in a mood to be amazed, it can amaze you that the system works at all. Amazement needn't be admiration : once you understand the system you may think there are better ones, or better ways to make this system work. I am only inviting you to reflect that whether this system works well or ill, in most countries and especially the countries with comparatively undirected economic systems, the system works the way ant colonies work.

Cet émerveillement ne peut qu'être accentué par le fait que fondamentalement, l'analogie que propose Schelling avec les sociétés d'insectes est loin de couvrir toutes les formes complexes d'organisation des sociétés humaines.

*Les actions médiatisées des insectes sociaux sont collectives et rigides, celles des chimpanzés sont au contraire individuelles et intelligentes - et celles des humains sont collectives et intelligentes*, écrivait Lestel (2001). Lorsque l'on observe le règne animal dans sa diversité, une chose est particulièrement frappante. Alors que la grande majorité des espèces vivent au sein de groupes relativement restreints - de l'ordre quelques milliers d'individus au plus - les espèces capables de s'organiser en grands groupes sociaux - dépassant les centaines de milliers voire les millions d'individus - se répartissent aux deux extrémités de l'échelle de la complexité individuelle.

En bas de cette échelle, se trouvent les insectes sociaux, fourmis, termites, abeilles, qui créent des communautés géantes pouvant atteindre plusieurs millions d'individus coordonnés de façon minutieuse. Ainsi, sont réalisées collectivement des tâches aussi étonnantes que la chasse collective, la culture de champignons, l'élevage d'autres insectes ou



la construction de véritables cathédrales de terre. Cela, sans qu'aucun des membres n'ait de représentation de l'activité globale de la société. Cette capacité d'une espèce à passer de comportements individuels relativement simples à des comportements collectifs complexes a été appelée intelligence collective [Bonabeau et collab. (1999); Bonabeau et Theraulaz (1994)].

En haut de l'échelle de la complexité, une seule l'espèce humaine donne naissance à des phénomènes de coordination et de coopération grande échelle dans une infinie mosaïque de cultures. Entre les insectes et nous, tout se passe comme si les individus des différentes espèces, en gagnant en autonomie et en intelligence, avaient perdu la capacité de se coordonner de manière efficace. Pour ces espèces intermédiaires, il semblerait que l'hétérogénéité inter-individuelle nuise à l'organisation des grands groupes sociaux.

L'intelligence collective des insectes sociaux commence à être bien comprise des éthologistes. Elle est due à des phénomènes émergeant de l'agrégation d'événements individuels. Ceux-ci sont possibles précisément parce que les membres de ces communautés ont des comportements stéréotypés, favorisés par la relative simplicité de leur organisme et leur grande proximité génétique.

La situation est toute autre du côté des sociétés humaines. L'hétérogénéité des individus à l'intérieur d'une même population, d'un même groupe social, est bien plus grande ; les différences inter-groupes également (voir par exemple les conclusions de l'étude sur plusieurs sociétés de Henrich et collab. (2005)). Par ailleurs, comme je l'ai défendu dans ma thèse [Chavalarias (2004)] puis dans Chavalarias (2007a) ou Chavalarias (2006b), les formes d'intelligence collective qui animent les sociétés humaines sont plus diverses que celles que l'on trouve chez les insectes sociaux, certaines étant totalement nouvelles dans l'évolution des espèces, nécessitant de fait d'autres approches de modélisation.

*Comment peut-on analyser la complexité des sociétés humaines et la diversité des cultures qu'elles développent dans le cadre d'une modélisation mathématique ou informatique ? Quelles sont les types de données qui nous permettront de tester ces modèles et en quoi les nouvelles sources de données, de type mega-données, peuvent-elles être constituer des opportunités pour développer de nouvelles approches de modélisation ? Voici le type de question qui est au cœur de mes recherches.*

## 2.1 Le triptyque de la modélisation en sciences humaines et sociales

Cette façon d'aborder le social peut sembler relativement générale par rapport à la manière dont il est habituellement étudié en sciences humaines et sociales. Les réponses attendues ont cependant de nombreuses implications sur le choix de nos approches de modélisation et dans notre compréhension des phénomènes ainsi appréhendés.

Dans l'introduction à son cours "Sciences Sociales et Sciences Cognitives" à l'École Polytechnique, Jean-Pierre Dupuy (1995) décrivait la scène des sciences sociales selon un diptyque correspondant en première approximation à l'opposition entre économie et sociologie :

L'économiste J. Duesenberry a eu un jour cette boutade : "l'économie n'est pas autre chose que l'étude de la façon dont les gens font des choix ; la sociologie n'est rien d'autre que l'étude de la façon dont ils s'arrangent pour

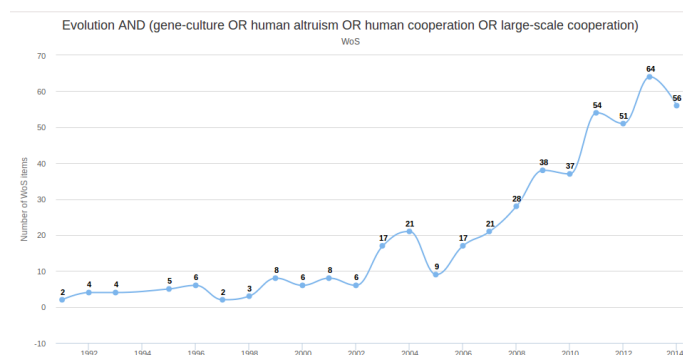


FIGURE 7 – Evolution du nombre de publications annuelles dans le Web of Science (WoS) pour la requête ("*gene-culture*" OR "*gene culture*" OR "*human altruism*" OR "*human cooperation*" OR "*large-scale cooperation*") AND (*evolution* OR *evolutionary*

ne pas avoir de choix à faire”. C’est une vision à peine caricaturale. Ainsi, le philosophe du choix rationnel, Jon Elster (1988), a-t-il pu écrire : “Une des lignes de clivage les plus tenaces à l’intérieur du domaine des sciences sociales est celle qui oppose deux formes de pensée que l’on associe respectivement aux noms d’Adam Smith et d’Émile Durkheim : c’est l’opposition entre *l’Homo œconomicus* et *l’Homo sociologicus*. Celui-là est sensé être guidé par une rationalité instrumentale, tandis que le comportement de celui-ci est dicté par des normes sociales. Le premier est “tiré” par la perspective d’avantages à venir alors que le second est “poussé” de derrière par des forces quasi inertielles (...). Il est facile de caricaturer le premier comme un atome asocial et autosuffisant et le dernier comme le jouet stupide des forces sociales”. Cette opposition se repère selon deux dimensions. L’une horizontale, est celle du temps, l’autre verticale, celle qui relie l’individu à la société.

Dans ce qui suit, j’ajouterai à ce diptyque un troisième tableau qui s’est constitué avec le développement des sciences cognitives et les théories évolutionnistes de la cognition. Il se déploie suivant une troisième dimension, l’évolution long terme de populations, et mobilise des concepts tels que “gene culture co-evolution”, “human altruism” ou “large-scale cooperation”. Encore relativement minoritaires dans l’ensemble des sciences sociales, la littérature est en croissance rapide (cf. fig. 7) et interdisciplinaire (cf. fig. 8). Ces recherches ont pour programme de mettre en évidence les déterminismes et les contraintes d’ordre biologique (au sens large) qui influent sur la constitution de l’ordre social, ainsi que les influences qu’a pu avoir en retour la culture sur l’évolution génétique. La communauté des systèmes complexes a contribué dès l’origine à ce courant de recherche, avec notamment de nombreux travaux du réseau de l’Institut des Systèmes Complexes de Santa Fe [Bowles et Gintis (2011); Boyd et Richerson (1988); Cavalli-Sforza et Feldman (1981); Fehr et Fischbacher (2003); Henrich et collab. (2005)].

Je n’adhérerai à aucune de ces trois tendances en totalité, mais chacune sera importante pour mettre en relief mes contributions. Je reprendrai en revanche des objets d’étude communs à l’ensemble de ces approches : l’origine de la diversité des cultures humaines, leur évolution et les méthodes pour les décrire.

## 2.2 Modèles et méga-données

("gene-culture" OR "gene culture" OR "human altruism" OR "human cooperation" OR "large-scale cooperation")  
AND (evolution OR evolutionary)

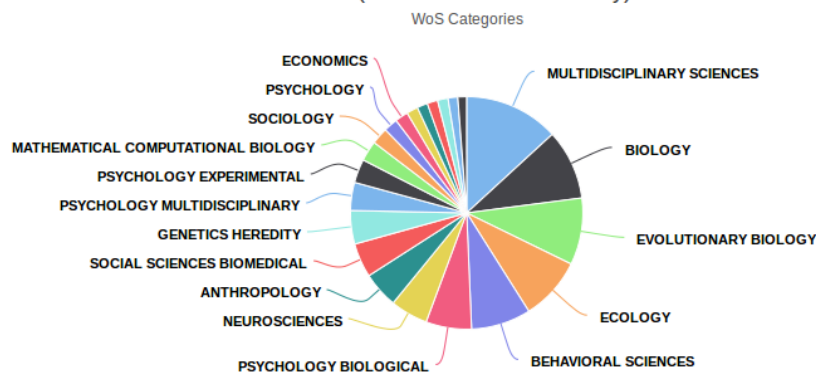


FIGURE 8 – Catégories de publication du WoS pour la requête ("*gene-culture*" OR "*gene culture*" OR "*human altruism*" OR "*human cooperation*" OR "*large-scale cooperation*") AND (*evolution* OR *evolutionary*). La forte interdisciplinarité de ces questions est directement reflétée par la diversité des catégories auxquelles elles sont rattachées.

Si la plupart des objets que nous allons aborder sont déjà anciens, leur approche par modélisation est relativement récente et a été favorisée par plusieurs facteurs :

- **Développement ou arrivée à maturité de cadres formels pour la modélisation en sciences sociales** : en parallèle au développement par l'économie, et dans une moindre mesure la sociologie, de la théorie des jeux (évolutionnaires) pour modéliser les dynamiques de population d'agents, d'autres communautés des sciences dites "dures" ont porté une attention grandissante à la modélisation en sciences sociales. Les mathématiciens ont apporté des modèles d'approximation dans un cadre continu de dynamiques de population en mobilisant les équations différentielles et les équations de réaction-diffusion (comme par exemple [Berestycki et collab. \(2015\)](#) ou [Hutson et Vickers \(2000\)](#)), les physiciens ont importé des modèles de champs moyen et de renormalisation dans ces interstices disciplinaires qui ont été appelés éconophysique et sociophysique, les informaticiens ont développés la simulation multi-agents pour étudier *in silico* les propriétés émergentes de collectifs d'agents artificiels et des scientifiques à l'interface math/infos ont appliqué la théorie des réseaux (dynamiques) à la compréhension des réseaux sociaux [[Chavalarias et Cointet \(2013a\)](#); [Palla et collab. \(2005\)](#); [Shahaf et collab. \(2013\)](#)], suivant en cela quelques précurseurs issus de la sociologie comme [Freeman \(1977\)](#); [Granovetter \(1978\)](#); [Schelling \(1978\)](#) qui ne formalisaient cependant que très peu les aspects dynamiques. Toutes ces contributions ont apporté des points de vue nouveaux sur la relation entre le collectif et les individus qui le composent,
- **Développement des sciences cognitives**. Les sciences cognitives sont essentielles pour identifier les capacités cognitives humaines susceptibles de donner lieu à des phénomènes d'auto-organisation au niveau du collectif (nouvelles formes d'apprentissage social, biais cognitifs, etc.). De manière heuristique, elles permettent d'identifier des capacités cognitives spécifiques à l'espèce humaine susceptibles de donner naissance à des manières de faire société qualitativement différentes des autres espèces animales (disciplines de la psychologie, l'anthropologie et l'éthologie cognitives). Celles-ci doivent d'une manière ou d'une autre être intégrées dans les entreprises de modélisation.

- **Avènement de l'ère des méga-données sociales, méthodes de reconstruction.** Le développement vigoureux de méthodes de reconstruction de dynamiques à partir des nombreuses traces numériques laissées par nos sociétés nous laissent entrevoir la diversité des conduites humaines et les motifs (au sens de *patterns*) de leurs évolutions avec une précision et sur des échelles de temps jusqu'alors inconcevables [Chavalarias et Cointet (2013a); Rudder (2014)]. En parallèle, nous assistons en sciences sociales au développement de nouvelles approches expérimentales qui tirent parti des nouvelles technologies pour étudier des cohortes de tailles rarement atteintes auparavant [Epstein et Robertson (2015); Salganik et collab. (2006)]. Ces deux aspects participent au phénomène des méga-données (*big data*) en science sociales. Ils offrent des opportunités inédites pour tester les modèles et ont un fort pouvoir heuristique pour la formulation de nouveaux modèles.
- **Développement exponentiel des moyens de traitement des données.** L'intégration de ces différents aspects (modèles, données, descriptions fines des individus) serait la plupart du temps impossible sans les nouvelles possibilités apportées par le calcul haute performance, soit dans le cadre de la résolution numérique d'équations mathématiques complexes, soit dans le cadre de simulations *in silico* de grands systèmes (souvent hétérogènes et avec une forte influence de la topologie des interactions).
- **Développement d'outils de visualisation et de dispositifs d'interaction avec les masses de données.** Proposée dès 1975 par de Rosnay (1975), les macroscopes sont maintenant une réalité [Börner (2011); Chavalarias et collab. (2015, 2011a)] au point que depuis 2015 une exposition leur est dédiée. Celle-ci sélectionnera pendant 10 ans les meilleures propositions de macroscopes<sup>5</sup>. Ces dispositifs, qui permettent de naviguer de manière interactive dans de larges ensembles structurés<sup>6</sup> de données permettent à la fois d'avoir une vue synthétique d'un phénomène collectif et d'en explorer les détails au niveau micro (entités et interactions). Ce sont des intermédiaires indispensables pour articuler les modèles en sciences sociales avec les données de terrains.

On peut d'ailleurs se faire une idée de la rapidité à laquelle se développent ces approches à partir d'une simple requête telle que (*sociology OR social*) AND ("*simulation*" OR "*game theory*" OR "*evolutionary game theory*" OR "*multi-agents*" OR "*agent-based*" OR MAS) dans une base telle que le Web of Science. Même si le corpus ainsi constitué ne représente que partiellement l'ensemble de la littérature sur le sujet, notamment en raison des biais de langue, des biais disciplinaires et de la polysémie de certains termes, il nous donne une première approximation de l'évolution de sa représentativité dans la portion littérature scientifique caractérisée par la requête *sociology OR social* (cf. figure 9). La pratique de la modélisation des dynamiques sociales est en plein essor depuis 2005, avec des temporalités relativement similaires à celle des recherches sur la modélisation de l'évolution culturelle observables sur la figure 7.

Insistons sur le fait que le phénomène des méga-données n'est pas neutre épistémologiquement et s'est accompagné d'une nouvelle forme d'empirisme, qui déclare la "fin de la théorie" [Kitchin (2014)]. Nous avons déjà souligné les limites cette approche dans la section 1.2 et, pour les raisons qui y sont exposées, je n'y adhère pas. Au contraire, mes recherches se rallient à l'idée que le phénomène des méga-données est l'occasion de

5. Le site *Places & Spaces Mapping Science*, <http://scimaps.org>

6. Par des modèles.

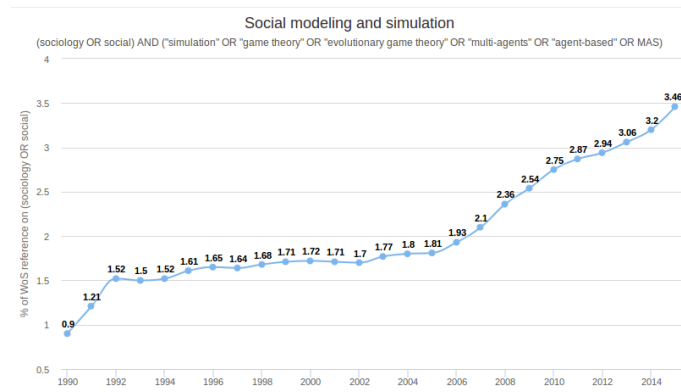


FIGURE 9 – Evolution du pourcentage de publications mentionnant les termes “simulation”, “game theory”, “evolutionary game theory”, “multi-agents”, “agent-based” ou “MAS” parmi les publications mentionnant “sociology” ou “social” dans le Web of Science. On remarquera la croissance très rapide à partir des années 2005.

renforcer les approches axées sur la théorie (*theory driven*) en les ancrant plus solidement dans les données, notamment pour ce qui est de la formulation des théories (*data-driven*). Ceci est une démarche commune à de nombreuses approches relevant des systèmes complexes, elle se démarque clairement des approches purement *fouille de données* (*data-mining*).

## 2.3 A propos de l'usage des termes *évolution culturelle* et *cognition sociale*

Les termes *évolution culturelle* et *cognition sociale* étant utilisés par plusieurs disciplines avec parfois des sens différents, précisons ce que recouvriront ces concepts dans ce qui suit.

### 2.3.1 Évolution culturelle

Deux types d'hérédité sont associés aux changements socio-culturels : l'évolution génétique et la transmission culturelle [Boyd et Richerson (1988)]. Je ne m'intéresserai dans ce qui suit qu'au second. Le terme *évolution culturelle* sera pris au sens de changements dans les comportements, les préférences, les valeurs ou les institutions d'une population sur des échelles de temps courtes par rapport aux temps biologiques (<100 ans). La perspective adoptée ici est d'ordre théorique et épistémologique : elle vise à dégager les principes structurants de l'évolution culturelle en amont des recherches modélisatrices proprement dites, puis d'en proposer des modèles.

Contrairement à des approches telles que la sociobiologie, qui voit dans le gène la source ultime de l'évolution culturelle, il s'agira au contraire d'identifier des moteurs de l'évolution culturelle sans invoquer une évolution par sélection naturelle d'un matériel génétique (ce qui contraint le type de modèles que nous aurons à utiliser). Dans cette approche, l'évolution génétique est neutre, voire même influencée par l'évolution culturelle plutôt que l'inverse (bien que sur des échelles de temps plus longues, des évolutions génétiques impliquant des changements phénotypiques majeurs aient inmanquablement de grandes chances d'interférer avec l'évolution culturelle).

La démarche adoptée se distingue de la plupart des approches formelles en économie par le fait qu'elle ne repose pas sur la notion d'agent optimisateur, notion qui sous-entend l'existence d'une mesure objective des utilités. Au contraire, la perspective choisie consiste à montrer de quelle manière un groupe définit le régime des utilités qu'il adopte par des critères internes, la valeur dérivant, de ce fait, des interactions entre les agents. On défendra donc une perspective dynamique sur l'évolution culturelle, impliquant la prise en compte de phases de constitution qui finissent par se stabiliser dans des formes instituées.

### 2.3.2 Cognition sociale

Le terme de cognition sociale est polysémique au sein de la communauté scientifique. La définition que nous considérerons dans les chapitres suivants est celle d'un traitement de l'information par l'ensemble des individus d'une société. Le traitement de l'information doit être entendu ici au sens de la cybernétique de second ordre [von Foerster (1981)], en prenant en compte les interprétations individuelles de toutes sortes auxquelles donne lieu ce traitement. L'un des objectifs de mes recherches est de formaliser les processus de cognition sociale comme un traitement d'information distribué sur un ensemble d'individus s'influçant mutuellement dans les objectifs qu'ils poursuivent et dans leurs manières d'être. Le résultat du processus de cognition sociale est alors l'émergence dynamique des identités et préférences des agents, ainsi que des groupes sociaux et institutions auxquels ils donnent naissance.

## 2.4 Résumé

*Mes recherches se situent à l'interface entre les humanités numériques, les sciences sociales computationnelles et les sciences cognitives. Elles portent sur l'analyse et la modélisation des dynamiques sociales, de la cognition sociale et de l'évolution culturelle avec comme objet de prédilection la dynamique et la structure des réseaux socio-sémantiques [Roth et Cointet (2010)].*

Depuis mon doctorat [Chavalarias (2004)], mes travaux ont porté sur l'étude des dynamiques sociales et de l'évolution culturelle. Après avoir modélisé les processus de différenciation sociale (valeurs, motivations, comportements) à partir de théories épistémologiques [Chavalarias et Gardelle (2008)] ou de découvertes en sciences cognitives, sociologie, anthropologie et éthologie [Chavalarias (2006b, 2007a, 2008c, 2009); Chavalarias et Gardelle (2008)], j'ai exploré les possibilités offertes par les humanités numériques pour développer des méthodes d'observation *in vivo* de dynamiques sociales ainsi que les conséquences et implications éthiques de la virtualisation des interactions sociales sur ces dynamiques [Chavalarias (2016)]. J'ai plus particulièrement étudié la dynamique des sciences [Chavalarias et Cointet (2008, 2013a); Chavalarias et Ioannidis (2010); Chavalarias et collab. (2016); Cointet et Chavalarias (2008)].

Il y a une dimension technologique incontournable dans les opportunités qu'apporte le phénomène des "méga-données" : de nouveaux dispositifs technologiques sont nécessaires en amont pour mettre en forme et traiter ces données et en aval pour tester les modèles (consistance analyse de sensibilité, prédictions). Les technologies mobiles ont quant à elles envahi de nombreux aspects du quotidien, ouvrant de nouvelles opportunités pour la collecte de données (crowdsourcing) et de nouvelles formes d'expérimentations en ligne multi-participants. C'est la raison pour laquelle j'ai investi une partie importante de mon activité dans le développement et la pérennisation de l'Institut des Systèmes Complexes de Paris Île-de-France (cf. section 7). Cet institut a catalysé de nombreux projets dans ce domaine (ANR, ERC, Equipex, etc.) où j'y ai constitué une équipe à même de faire l'interface entre SHS et cette dimension technologique. J'ai également été attentif aux enjeux de transfert et de valorisation (cf. section 6).

*Interdisciplinaires par nature, mes recherches relèvent de plusieurs axes complémentaires [sections CNRS concernées] :*

- **Humanités numériques [53, 06]** : web-mining, fouille de données et reconstruction de dynamiques sociales.
- **Épistémologie et méthodologie scientifiques [53, 35]** : méthodes de reconstruction de dynamiques des champs scientifiques à partir des méga-données, modélisation de la dynamique et structure des disciplines scientifiques,
- **Économie cognitive et en sociologie [37, 06]** : modélisation de la cognition sociale et évolution culturelle, théorie des jeux évolutionnistes et dynamiques mimétiques,
- **Environnements informatiques pour l'apprentissage humain [06]** visualisation interactive de données, nouvelles interfaces pour les humanités numériques.

Ces recherches se sont accompagnées d'une réflexion épistémologique sur le développement des approches "systèmes complexes" [Bourgine et collab. (2008); Deffuant et collab. (2015)] et des technologies de l'information et de la communication [Chavalarias (2016)] (cf. notamment section 1).

### 3 Reconstruction des dynamiques sociales à l'ère des méga-données

En moins d'une génération, les nouvelles technologies ont profondément transformé nos pratiques quotidiennes dans quasiment tous les domaines, entraînant un changement radical des possibilités d'analyse des activités sociales et économiques via leur traces numériques. Entre 1995 et 2014<sup>7</sup>, la couverture d'Internet est passée de 0,6% à 39% de la population mondiale, la couverture de téléphonie mobile de 1% à 73%. Aux États Unis, le temps moyen passé par la population adulte sur des médias numériques a plus que doublé depuis 2008, pour atteindre 5,6 heures par jour. Plus de 80% des 12-24 ans utilisent les médias sociaux.

Sont déjà analysés, à des fins scientifiques, commerciales ou stratégiques (cf. Chavalarias (2016)), des corpus numériques issus du milieu académique, de la presse et des la blogosphère, des bases de dépôt de brevets, des réseaux sociaux en ligne (science, messageries instantanées, loisirs, rencontres, etc.), des plates-formes économiques (achat/vente de biens divers, services et loisirs, bourse en ligne, marchés prédictifs, etc.), des plate-formes collaboratives (Wikipedia, Digg, etc.). Cette liste est loin d'être exhaustive.

Outre le fait que ces nouvelles données nous permettent d'étudier des domaines qui

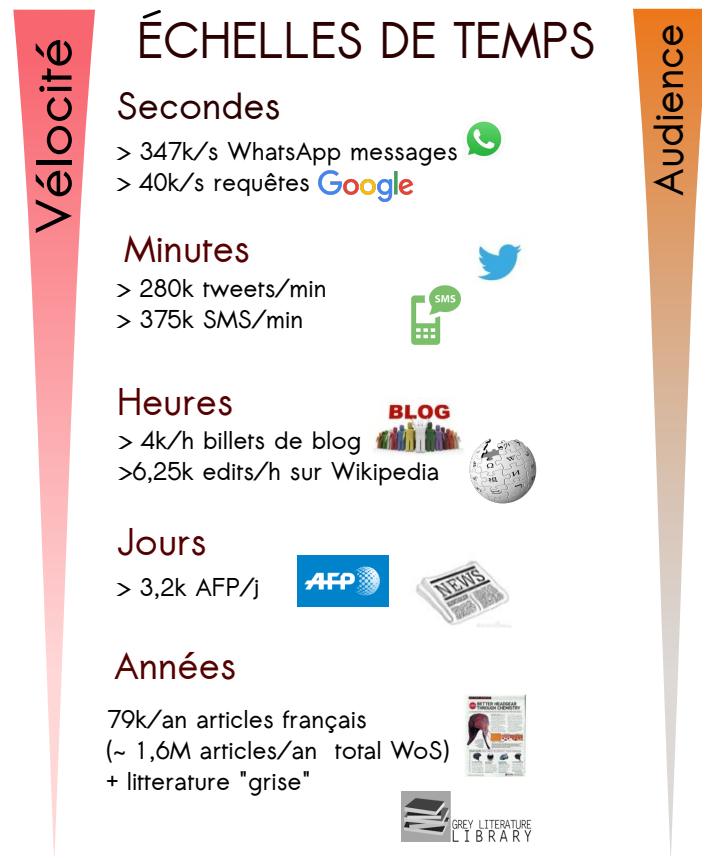


FIGURE 10 – Exemples de média pouvant être support d'une reconstruction phylométique. Chaque canal de production a ses temporalités propres qui sont fonction à la fois du type de contenu, des types de communautés qui les produisent et de leur audience. Un thème comme le changement climatique se retrouvera dans l'ensemble des contenus mentionnés ici. D'autres thèmes pourront être confinés à un média particulier comme par exemple la technique du "point de croix" (tricot), abondamment discutée dans une partie spécifique de la blogosphère.

7. cf. Meeker (2015)



ne nous étaient pas accessibles ou pour lesquels nous n'avions que des données partielles, l'étendue des domaines concernés par la révolution des méga-données et la très fine granularité des données (on peut savoir jusqu'au temps passé pour chaque action d'une activité donnée) offre la possibilité d'étudier diverses arènes de production des connaissances avec des méthodologies transversales et de comparer les résultats obtenus de manière multi-échelle en extension et en temps. J'ai commencé à mettre en œuvre cette approche inter-corpus dans le projet *Tweetoscope Climatique* [Chavalarias et collab. (2015)] (cf. 3.4) où ont été collectés 20 millions de Tweets et 3 millions de pages Web associées, qui ont ensuite été analysés avec 58 millions d'articles du Web of Science. Ce projet m'a permis de développer une méthodologie et des technologies pour la collecte de données de manière consistante, dans différents types de médias (science, blog, presse et réseaux sociaux). Ces méthodologies et technologies constituent une base indispensable à mon projet de recherche actuel.

La science est l'une des productions culturelles les mieux structurées. C'est aussi l'un des premiers domaines d'activité à avoir vu sa production presque intégralement numérisée. Elle constitue un objet d'étude unique pour l'analyse et la compréhension des dynamiques culturelles (motifs d'innovations, dynamiques collectives, morphogenèse des réseaux socio-sémantiques, processus de différenciation, etc.) à partir de traces numériques. Elle est un objet d'étude privilégié de mes recherches. La figure 11 résume les types de modélisation couverts par mes recherches dans ce domaine : reconstruction phénoménologique, reconstruction théorique et interfaces pour l'exploration interactive des résultats des reconstructions.

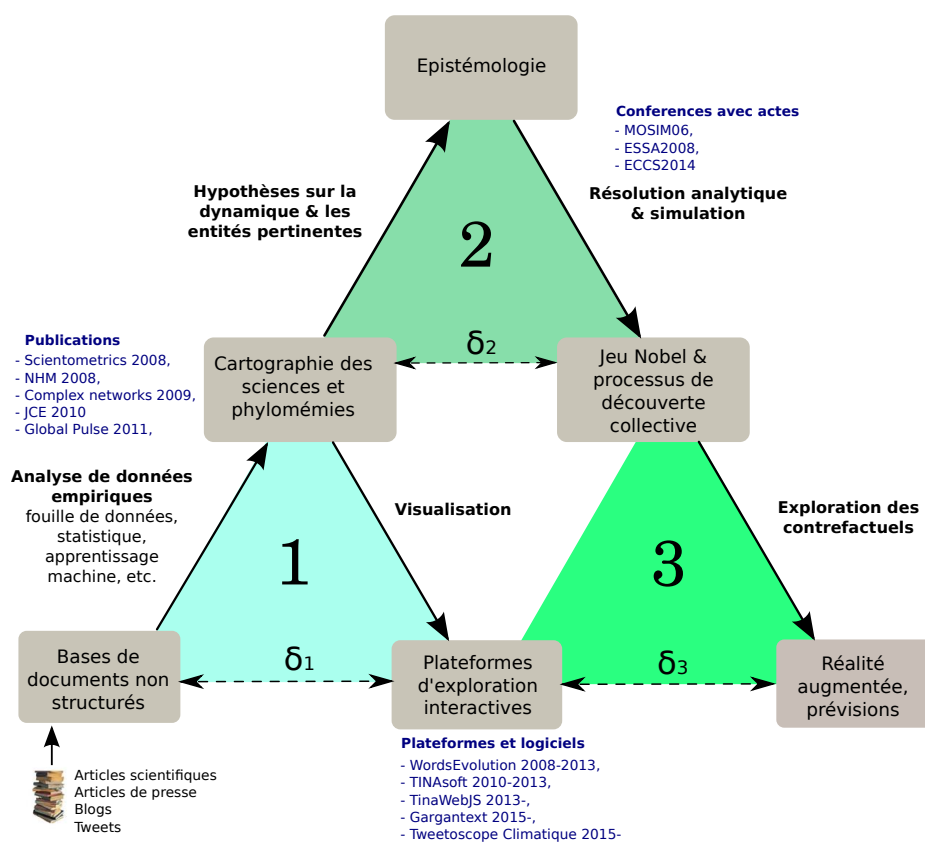


FIGURE 11 – Types de modélisation couverts par mes recherches sur dynamique et la structure des disciplines scientifiques.

## 3.1 La thèse de Popper est-elle réfutable ?

*We don't know, we can only conjecture.* Popper, 1934

### 3.1.1 La philosophie des sciences comme objet de modélisation

La première question de modélisation que je me suis posée quand je me suis intéressé aux systèmes sociaux portait sur la modélisation de la science elle-même, en tant que processus d'élaboration collective de connaissances.

L'histoire, la philosophie et la sociologie des sciences sont des branches de la science bien établies, dont la longue histoire et les interactions avec les sciences dites "dures" ont façonné notre manière de concevoir le développement de la connaissance, et par conséquent les outils et méthodologies adoptés par la *science normale*.

Pourtant, la plupart des personnes s'engageant dans un cursus universitaire dit "scientifique" n'ont bien souvent qu'une connaissance partielle de ces domaines, empreinte principalement des visions galiléenne et cartésienne de la science.

Réciproquement, alors que de nombreux philosophes ont théorisé le développement des connaissances scientifiques (Kuhn (1962), Popper (2002, 2006), Hull (1988) ou Darnemark (2003) pour n'en citer que quelques uns) il n'y avait pas, jusqu'à très récemment, de modélisation de ces théories.

Les premiers rapprochements se sont opérés dans les années 1960 et 1970, lorsque les traces laissées par la science (articles, biographies, statistiques administratives, etc.) ont été estimées suffisamment exploitables pour donner prise à une approche quantitative. Elles ont donné lieu à deux types de méthodes distinctes : "*description-prédiction*" et "*hypothèses-tests*"<sup>8</sup> Gilbert et Woolgar (1974). La première part de données disponibles et cherche à les décrire avec une fonction mathématique. Une interpolation de ces données peut alors permettre d'émettre des prédictions. La seconde fait le chemin inverse. Elle commence par faire des hypothèses sur les processus sociaux à l'œuvre en science. Les implications de ces hypothèses sont alors explorées pour résulter bien souvent dans la proposition d'une fonction mathématique qui les synthétise et produit un ensemble de caractéristiques empiriquement vérifiables découlant de ces hypothèses. Ces dernières sont ensuite vérifiées ou rejetées suivant que les données corroborent ou pas ces caractéristiques empiriques.

Les premières études se sont attachées à la mesure de la croissance des communautés scientifiques, les motifs des citations.

La question fondamentale de l'articulation entre l'activité de recherche privée, la publication puis l'adoption ou non de cette recherche par la communauté scientifique n'a pas été immédiatement abordée par les modélisateurs alors que cela aurait dû être un sujet de prédilection, la modélisation étant un très bon outil pour effectuer le passage du local au global, de l'activité individuelle à l'activité collective.

### 3.1.2 Nobel, le jeu de la découverte collective.

En 1998, je me suis inspiré de l'épistémologie de Karl Popper [Popper (2006)] pour proposer un modèle de la dynamique de publication/réfutation d'une communauté en fonction de la manière dont ces activités sont reconnues par les institutions et la communauté scientifique. Nous connaissons tous le dicton "*publish or perish*" qui souligne

8. Certains les appellent aujourd'hui ces méthodes *data-driven* et *theory-driven*.

la dérive productiviste des systèmes scientifiques contemporains. Au-delà de cette caricature, il y a de nombreuses nuances dans les systèmes de production de la science.

L'activité de publication est mise en valeur de différentes manières en fonction des pays, des types de science et des types d'établissement. Dans certains cas, les enseignants-chercheurs "publiant", peuvent obtenir des décharges de cours, dans d'autres, un nombre minimum de publications par an est nécessaire pour obtenir un poste ou le conserver, ailleurs les publications augmentent les chances d'obtenir une prime de recherche (ex. PEDR) ou font baisser les frais de scolarité pour les étudiants, etc.

Des questions similaires se posent sur la réfutation. Quel est le coût d'une réfutation pour celui qui est réfuté ? Comment sont reconnues la réfutation ou la reproduction des travaux des pairs ? Dans certains pays et certains domaines, les auteurs de faux résultats sont mis au ban de leur communauté, ailleurs la punition est moins sévère.

Prenons le cas de la France avec *l'affaire d'Olivier Voinnet*. Ce chercheur de renommée internationale, membre de l'académie des sciences et médaille d'or de l'*European Molecular Biology Organisation* a publié entre 1998 et 2015 au moins 20 articles dont on a découvert en 2015 qu'ils comportaient de faux résultats et des manipulations de données. Les sanctions à son encontre ont été très différentes entre le CNRS (France) d'où il a été exclu pour 2 ans, et l'ETHZ (Suisse), où il dirige un laboratoire, qui a été plus clémente. Ce "manquement caractérisé aux règles et aux bonnes pratiques de présentation des données scientifiques dans treize articles" serait-il passé à travers les mailles du filet si les nombreux scientifiques qui ont cité ses articles avaient pris soin de vérifier ses résultats ? Pire, une relectrice anonyme des articles d'Olivier Voinnet avait fait part de graves soupçons de fraude dès 2003, auprès de la revue qui l'avait missionnée comme relectrice. Ce signalement n'avait eu aucun effet sur la publication de l'article, ce qui montre à quel point la réfutation est prise au sérieux comme activité scientifique.

Ce type d'affaire révèle avant tout une dissymétrie entre l'importance accordée à la publication et à la réfutation dans le système de la recherche, dont il convient d'étudier les effets. La montée en puissance du principe de reproductibilité<sup>9</sup> ou la dénonciation du nombre de faux résultats en science [Ioannidis (2005)] laisse entrevoir un changement de mentalités, mais dans quelle mesure ?

Ces différents modes de valorisation de l'activité scientifique se traduisent très bien dans le cadre de la théorie des jeux. Lorsque l'on prend comme référence l'épistémologie de Popper et le principe suivant lequel on ne peut pas démontrer la vérité d'une théorie, la sélection des théories par conjectures et réfutations met au centre du jeu la question du temps, des anticipations croisées sur les pratiques de chacun et de la reconnaissance des différents types d'activité par les institutions ayant une influence sur la carrière des chercheurs.

C'est ce que j'ai traduit à travers le modèle appelé *jeu Nobel* [Chavalarias, D. (1998)], qui modélise certains aspects de la théorie de Popper de manière à en dériver des prédictions empiriquement vérifiables.

Ces travaux ont fait partie des toutes premières propositions de modélisation de la dynamique des sciences dans un cadre multi-agents (avec [Gilbert, N. (1997)]), domaine de recherche qui est actuellement en pleine effervescence [Edmond, B. et collab. (2011); Payette (2012)]. Une première approximation analytique de la dynamique du modèle a d'abord démontré l'existence de cycles dans les processus de découverte collective, cycle dont les temps caractéristiques dépendent des manières dont les activités de publication et de réfutation sont reconnues socialement. J'ai confirmé par la suite ces résultats par une

9. Voir le *Special* de la revue *Nature* sur le sujet <http://www.nature.com/news/reproducibility-1.17552>

approche multi-agents [Chavalarias (forthcoming); Chavalarias et collab. (2006); Chavalarias et Gardelle (2008)].

J'ai mis en évidence plusieurs autres faits stylisés concernant la dynamique des sciences : effets de la taille de la communauté scientifique, des incitations à la publication, de la valorisation des réfutations, de la difficulté du champs et de l'abondance des objets de recherche sur la *qualité* et la *vitesse* du processus de découverte collective, deux caractéristiques qui doivent faire l'objet d'un compromis individuel, qui se transpose en un compromis collectif (cf. fig. 12 pour un exemple de l'effet de l'incitation à la publication).

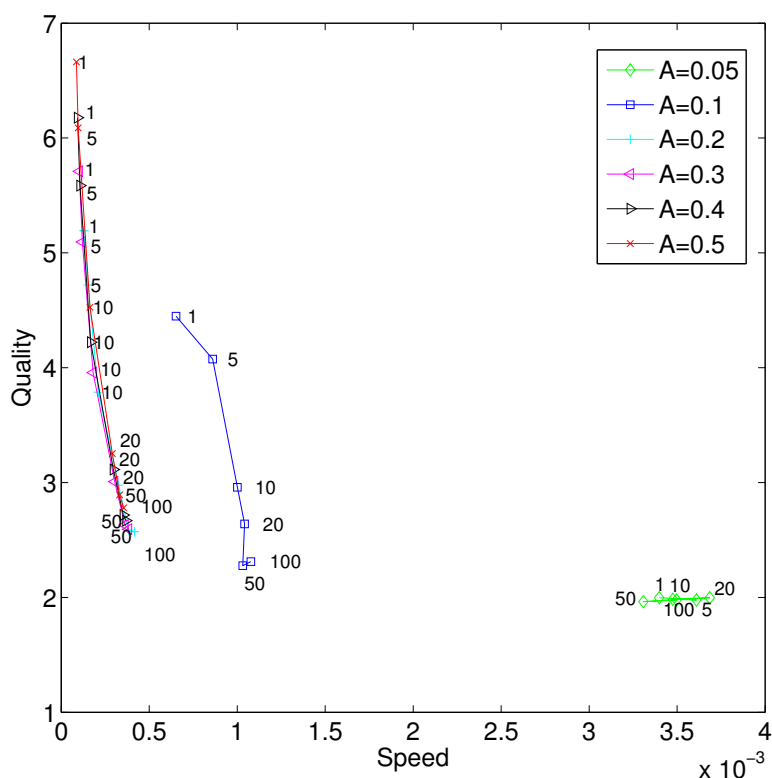


FIGURE 12 – **Simulation de la dynamique d'un jeu Nobel.** Pour une population de 100 agents explorant un espace de théories de différentes qualités, diagramme de dépendance de la qualité des publications et de la vitesse du processus de découverte en fonction de l'incitation à la publication (sur une échelle de 1 à 100, indiquée à côté des points de simulation). Chaque point est la moyenne de 100 simulations à différents niveaux d'avancement  $A$  du processus de découverte (le niveau  $A$  est indiqué par la légende.  $A = 0$  au début du processus et  $A = 1$  lorsque toutes les théories valides ont été publiées). Ce diagramme montre clairement le compromis existant entre vitesse du processus de découverte et sa qualité. A noter également qu'au début d'un processus de découverte, l'incitation à la publication a très peu d'impact sur la qualité ou la vitesse alors qu'ensuite, elle réduit principalement la qualité sans accroître significativement la vitesse. *Définitions.* Qualité :  $-\log_{10}$  de la probabilité pour d'une théorie publiée prise au hasard passe un test de réfutation, moyennée sur l'ensemble du processus de découverte. Vitesse : nombre moyen de bonnes théories découvertes par chercheur.

J'ai toujours eu l'ambition de tester expérimentalement ou empiriquement les prédictions de ce modèle. Mon intérêt pour la scientométrie est issu en partie de la volonté de corroborer ou réfuter empiriquement les prédictions de mes modèles par des observations *in vivo* (voir notamment 3.3). La première tentative, en 2005 [Chavalarias et collab.

(2006)], de les tester expérimentalement est restée inachevée, essentiellement à cause de la difficulté de mise en œuvre d'expériences multi-joueurs à cette époque. Le développement des technologies de l'information m'a permis récemment de relancer ce programme de recherche et j'ai pu, dans le cadre du projet *Science en Poche*, dont je suis le coordinateur, développer avec Maziyar Panahi (ISC-PIF) et Jean-Baptiste Leenart (ISC-PIF), une version en ligne et multi-joueurs du jeu Nobel (<http://nobelgame.org>). Des expériences préliminaires et prometteuses ont déjà été effectuées depuis 2014, notamment lors de la conférence ECCS14 où une trentaine de joueurs ont participé pendant 3 jours à plusieurs expériences.

Ces recherches proposent une manière de penser l'impact des politiques de la recherche sur le jeu social qui régule les activités de conjecture et de réfutation et détermine, *in fine*, la qualité de la science. Comme pour le cas du modèle de Schelling (cf. 1.2), on ne peut prétendre que ce modèle reflète ce qui se passe effectivement dans le milieu de la recherche, mais pour autant qu'on puisse démontrer que les politiques d'incitations à la publication (et la réfutation) aient des effets sur le comportement des chercheurs, alors il est concevable que des phénomènes similaires à ceux observés dans le jeu Nobel soient à l'œuvre dans la régulation de la production scientifique.

L'une des critiques connues de l'approche popperienne de l'évolution des sciences, qui pose l'activité de conjectures et réfutations comme démarcation de l'activité scientifique, est qu'elle n'est pas elle-même réfutable. Mes travaux sont aussi une proposition pour répondre à cette critique, en démontrant que l'on peut dériver des prédictions empiriquement vérifiables à partir d'hypothèses sur la nature des théories scientifiques et les processus qui permettent de les sélectionner.

#### PUBLICATIONS ET TRAVAUX ASSOCIÉS

- Chavalarias, D., (a paraître). What's wrong with Science ? Modeling the collective discovery processes with the Nobel Game, *Scientometrics*.
- Chavalarias, D., 1998. La thèse de Popper est-elle réfutable (Mémoire de Master 2). École Polytechnique, Paris.
- Chavalarias, D., Charron, S., Gardelle, R., Bourgin, P., 2006. NOBEL, Le jeu de la découverte scientifique, Une approche analytique, expérimentale et computationnelle, in : Conférence Modélisation, Optimisation et Simulation Des Systèmes, Défis et Opportunités (MOSIM'06). Presented at the conférence Modélisation, Optimisation et Simulation des Systèmes, Défis et Opportunités (MOSIM'06), Rabbat.
- Chavalarias, D., Gardelle, V., 2008. Social simulation of collective a discovery process : the Nobel Game, in : 5th Conference of the European Social Simulation Association. 5th conference of the European Social Simulation Association, Brescia, Italy.

## 3.2 Reconstruction de la dynamique des sciences

*Mais comme dans les cartes générales du globe que nous habitons, les objets sont plus ou moins rapprochés, et présentent un coup d'œil différent selon le point de vue où l'œil est placé par le Géographe qui construit la carte, de même la forme de l'arbre encyclopédique dépendra du point de vue où l'on se mettra pour envisager l'univers littéraire. On peut donc imaginer autant de systèmes différents de la connoissance humaine, que de Mappemondes de différentes projections ; et chacun de ces systèmes pourra même avoir, à l'exclusion des autres, quelque avantage particulier.*  
*d'Alembert (1751)*

Deux raisons m'ont amené à l'étude de la dynamique des sciences à partir de l'analyse à grande échelle de ses productions. Ma première motivation était de trouver des données susceptibles de corroborer ou d'infirmer les prédictions relatives à la modélisation que j'avais faite de l'épistémologie de Karl Popper [Chavalarias et Gardelle (2008); Chavalarias, D. (1998)] (cf. 3.1).

La seconde tient du nomadisme de concepts. J'ai côtoyé à l'Institut des Systèmes Complexes de Paris Île-de-France une équipe de biologistes du développement et de modélisateurs menée par Nadine Peyrieras et Paul Bourguine. Ils travaillaient à la reconstruction *in silico* du linéage cellulaire d'animaux modèles à partir d'images issues de nouvelles techniques de microscopie confocale [Olivier et collab. (2010)]. Ils étaient capables de suivre la position de toutes les cellules et leurs divisions à travers le temps à partir de quelques cellules souches jusqu'à l'organisme constitué. Des outils de visualisation leur permettait de retracer le processus de formation des organes en naviguant dans un embryon reconstitué *in silico* en 4D (3D + le temps, cf. figure 13).

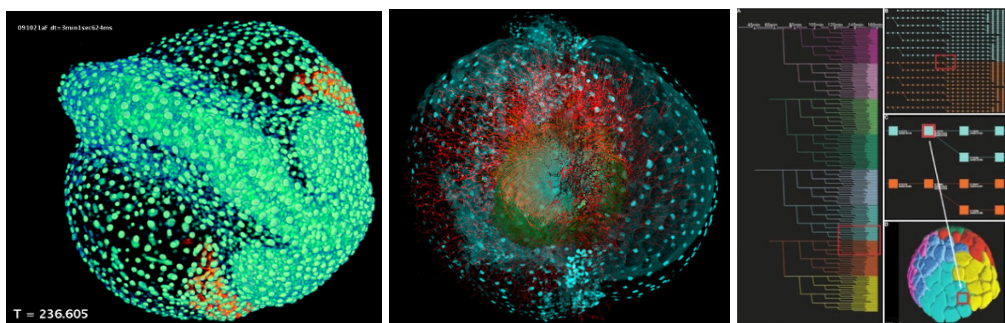


FIGURE 13 – Trois images de reconstruction de l'embryogenèse d'un zebrafish. a) reconstitution 3D à partir des images 2D de microscope ; b) suivi des trajectoires cellulaires ; c) reconstruction du linéage cellulaire. Source : CNRS, Projet Bioemergences, Nadine Peyriéras.

Cette recherche sur la morphogenèse en embryologie a certainement été aussi inspirante pour moi qu'a pu l'être en son temps la recherche sur l'évolution des espèces pour les pionniers de l'épistémologie évolutionniste. La question de savoir s'il était possible

de reconstruire l'évolution des sciences de manière à "voir" et quantifier l'évolution des disciplines et à interagir avec cette structure multi-échelle ainsi reconstruite ne m'a plus quitté.

### 3.2.1 Les analyses par co-citations et par co-termes

L'analyse quantitative des sciences (et dans un second temps de leur évolution) est un domaine dont on peut retracer la naissance aux années 1950 [Garfield (1955)]. Elle a commencé avec l'étude des motifs de citation qui a mobilisé principalement deux méthodes : le *bibliographic coupling* qui s'intéresse à la manière dont deux publications peuvent comporter des références communes [Kessler (1963)], et l'analyse par co-citation (*co-citation analysis*, Garfield (1974)), qui étudie la manière dont les articles sont cités conjointement. Ces méthodes étaient initialement tournées vers la question de la recherche d'information et de l'identification des articles pertinents pour une requête donnée. Elles n'avaient pas pour but d'aboutir à des conclusions générales sur la structure de la science et son évolution.

Une deuxième branche de l'analyse quantitative s'est développée autour de l'analyse des co-occurrence de termes dans les documents ou analyse par *co-termes* (*co-word analysis*, Callon et collab. (1983)) inventée Centre de Sociologie de l'Innovation CNRS-Mines de Paris. Cette approche était explicitement tournée vers l'analyse des dynamiques des sciences dans une perspective sociologique, pour laquelle l'analyse par co-citation était considérée comme insuffisante et limitée. Outre les points faibles déjà bien connus de l'analyse par co-citation (par ex. faible sensibilité - ou *recall* - de l'identification des domaines récents, fragmentation de certaines communautés identifiées [Braam et collab. (1991)]), Callon et collab. (1983) avançaient le fait que la citation en tant que pratique sociale n'est pas bien définie et peut recouvrir des significations aussi diverses que l'allégeance, la reconnaissance, la réciprocité, etc. Ceci rend délicate l'interprétation des structures cognitives que cette méthodologie met en évidence.

Par ailleurs, comme l'ont fait remarquer ces mêmes auteurs, la pratique de la citation est absente de certains contextes de production de connaissance et, dans les contextes où elle est présente, peut avoir des significations très différentes. L'analyse par citations n'est donc pas pertinente pour une étude générique des contextes de production scientifique, que Latour et Woolgar (1986) ont proposé de regrouper sous le concept de *literary inscription*, et qui englobe les articles scientifiques, mais aussi les rapports, les projets, les brevets, etc.

Partageant ce point de vue, et ayant également pour objectif d'étendre l'analyse de la production collective de connaissances à l'ensemble des contenus textuels du Web (blogs, articles de presse, tweets, etc.) mes recherches se sont naturellement inscrites dans le cadre de l'analyse par co-words plutôt que par citations. Bien que ces approches méritent d'être mises en complémentarité, je me concentrerai dans ce qui suit sur les recherches concernant l'analyse par co-termes.

Les approches quantitatives de l'évolution des sciences se sont considérablement développées ces dernières années grâce à l'arrivée à maturité de la numérisation des productions scientifiques autant en terme de volume annuel qu'en terme de nombre d'années couvertes (cf. la figure 14 pour une estimation de la dynamique).

Ces analyses quantitatives ont été menées dans avec divers objectifs. Aux côtés de l'objectif initial qui était la recherche d'information, la question de la cartographie des

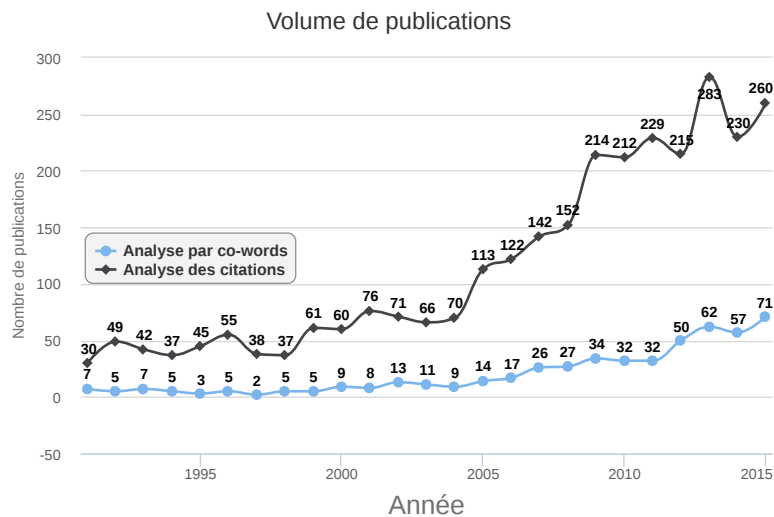


FIGURE 14 – Evolution du nombre de publications mentionnant “(co-word OR “co-occurrence analysis”)” et “(“co-citation” OR “bibliographic coupling” OR “citation analysis”)” dans leur titre ou résumé dans le WoS.

sciences (*science maps*) a pris une importance grandissante (voir [Borner et collab. \(2003\)](#) pour un état de l’art).

Jusqu’à la fin des années 2000, la grande majorité des travaux dans ce domaine ont eu pour objectif d’établir *une* carte d’un domaine de la science à *un* moment donné<sup>10</sup>, la dynamique pouvant être traitée par la juxtaposition d’analyses de cartes statiques à différentes périodes (par ex. [Peters et van Raan \(1993\)](#)).

Mes travaux ont eu, dès 2006, des objectifs très différents : analyser de manière multi-échelle l’évolution des sciences dans son ensemble et produire des dispositifs interactifs permettant de se faire une représentation individuelle ou collective de tout ou partie de la science.

Un très grand nombre de dimensions est nécessaire pour appréhender les différents aspects de la science (disciplinaire vs. interdisciplinaire, théorique vs. expérimentale ou empirique, émergente vs. mainstream, orthodoxe vs. hétérodoxe, dimension méthodologique, dimension technologique, etc. ). Toute opération de synthèse ou de représentation dans le cadre d’une représentation réduira nécessairement l’espace des représentations possibles, privilégiant un point de vue particulier. L’activité de cartographie d’un ensemble de documents est une activité de co-construction médiatisée par l’outil logiciel, contrainte autant par la structure cognitive des auteurs des documents analysés que par la structure cognitive de la ou des personnes qui établissent la carte. LA carte, comme certains la rêvent, n’existe pas. Pour explorer la structure de la science, nous avons donc besoin de dispositifs nous permettant de changer de point de vue aussi souvent que nécessaire pour cerner ces particularités.

La question de l’interaction avec les visualisations des sciences (et donc de produire des dispositifs permettant cette interaction) est loin d’être un détail. Elle contraint en effet le type de structure que l’on peut communiquer à travers la recherche et donc chercher à

10. Voir l’exposition *Places & Spaces* organisée par le *Cyberinfrastructure for Network Science Center* (Indiana, US.) pour une sélection de visualisations sur ces dix dernières années, [http://sci-maps.org/browse\\_maps.html](http://sci-maps.org/browse_maps.html)



mettre en évidence. Jusqu'à très récemment, très peu de logiciels ont été proposés pour naviguer de manière multi-échelle dans les cartes de sciences et c'est probablement pour cette raison que très peu de travaux questionnent le caractère multi-échelle de la science. Mes publications sont quasiment les seules dans le domaine de l'analyse par co-termes à faire état de cette question.

La visualisation et l'interaction avec les reconstructions comporte un autres aspect fondamental, celui de la validation externe. Lorsque j'ai effectué mes premières recherches sur la reconstruction de la phylométrie des sciences [Chavalarias et Cointet (2009)], nous avons de très belles "régularités" et les résultats statistiques étaient convaincants. Mais quelle proportion des branches de phylométries faisaient vraiment sens ? L'étude Chavalarias et collab. (2010) que nous avons faite pour le MESR pour qui comprenait la réalisation d'une plate-forme de navigation dans les phylométries de la blogosphère a été une expérience déterminante pour améliorer notre approche là où elle produisait encore des résultats aberrants. Dans ce domaine, rien ne vaut l'utilisation des résultats de recherche en situation de production par des utilisateurs finaux pour tester leur robustesse.

Mes principaux apports dans ce domaine concernent donc :

- **La reconstruction multi-échelle des domaines scientifiques.** [Chavalarias et Cointet (2008); Cointet et Chavalarias (2008)] Avec notamment l'utilisation et le développement d'algorithmes de détection de communautés avec recouvrement. Ce caractère est essentiel pour prendre en compte le caractère potentiellement polysémique des termes. Je me suis notamment intéressé à la modélisation des relations générique/spécifique entre les termes et à la question de l'identification *bottom-up* de structures meso et macro prenant en compte ce type de relation.
- **La reconstruction de la dynamique des sciences** . L'aspect temporel a fait partie des le départ de mon programme de recherche et c'est dans de domaine que j'ai apporté les contributions les plus originales (Chavalarias et Cointet (2013a,b)). Voir la section 3.2.3.
- **Le développement de nouvelles approches de visualisation et d'interaction avec les données** J'ai toujours été convaincu que la manipulation des reconstructions de la structure de la science faisait partie intégrante du mécanisme d'appropriation du contenu qu'elle véhicule. J'ai donc développé des approches permettant aux utilisateurs d'interagir avec ces reconstruction, voire de les générer eux-mêmes et les manipuler de manière interactive et multi-échelle (de la vue haut-niveau jusqu'au unités textuelles et sémantiques.). Voir section 6.

### 3.2.2 Phylométrie de sciences

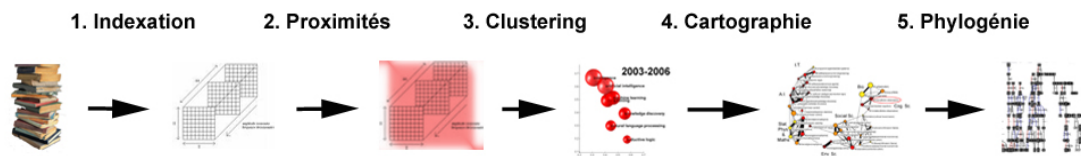


FIGURE 15 – Étapes pour la reconstruction d'une phylométrie des sciences.

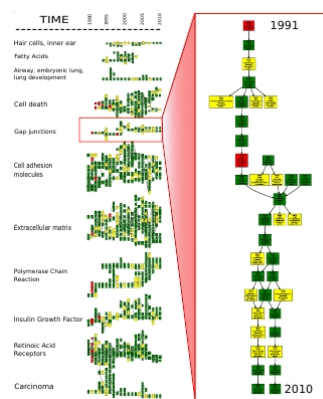


FIGURE 16 – Extrait de la phylométrie du domaine de l'embryologie.

méthodologie pour la reconstruction de la dynamique des sciences à partir d'une base de documents non structurée, en testant la robustesse via des analyses de sensibilité et la reconstruction sur plusieurs jeux de données ainsi que des régularités dans l'évolution des sciences. Entre le pre-print et la publication, nous avons développé la partie amont de text-mining qui a permis de tester la généralité et la robustesse de nos résultats en dehors de toute intervention humaine.

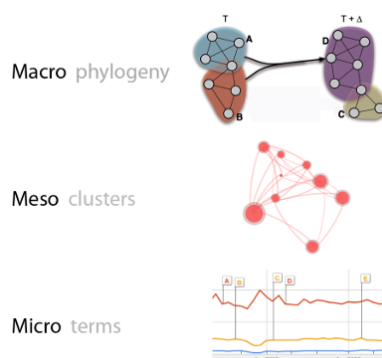


FIGURE 17 – Échelles d'analyse de la dynamique de l'aspect thématique des champs scientifiques : dynamique micro des termes, meso des champs, macro des domaines.

En mobilisant des méthodes issues des systèmes complexes [Palla et collab. (2007)] et du TAL appliquées à l'analyse des grandes archives documentaires et bibliographiques (Medline, « Web of Science », etc.) j'ai effectué à partir de 2006 des recherches sur la reconstruction phénoménologique de la dynamique des sciences. Dans Chavalarias et Cointet (2013a) (pre-print Chavalarias et Cointet (2009)), j'ai proposé le concept de *phylométrie* pour décrire la structure de cette évolution et un workflow (fig. 27) pour cette reconstruction que j'ai implémenté sous Matlab. Ces travaux proposent une

J'ai proposé une typologie des nœuds d'une phylométrie (fig. 24) dont l'analyse a permis de confirmer certaines intuitions de Callon et collab. (1991a) : les variations d'indicateurs de la cohérence d'un champ (par exemple la *densité*), qui peuvent être calculés à partir de valeurs "instantanées" sans connaître la structure de l'évolution des sciences sont corrélées avec la structure globale de la phylométrie (cf. fig. 19).

Les branches des phylométries exhibent de nombreuses particularités morphologiques qui semblent refléter les types de régimes de la recherche qui s'y déroulent. La caractérisation fine et inter-domaines de ces typologies ouvre des

perspectives de comparaison avec les régimes de sciences prédits par des auteurs comme [Kuhn \(1962\)](#), [Popper \(2002, 2006\)](#), [Hull \(1988\)](#) ou [Danermark \(2003\)](#). Ce projet fait partie de mon programme de recherche actuel (cf. partie 5 et je co-dirige actuellement une thèse dont l'objectif est de produire des reconstructions à grande échelle et inter-domaines des phylomémies, ce qui comporte en soi un certain nombre de défis théoriques et technologiques.

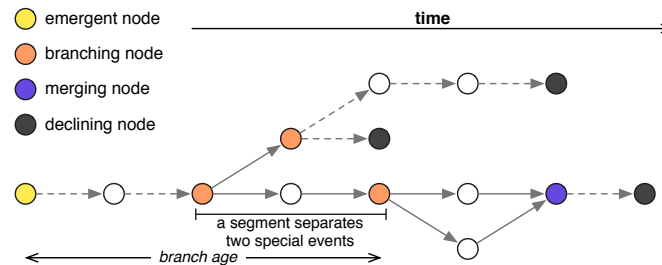


FIGURE 18 – Typologie des nœuds d'une phylométrie.

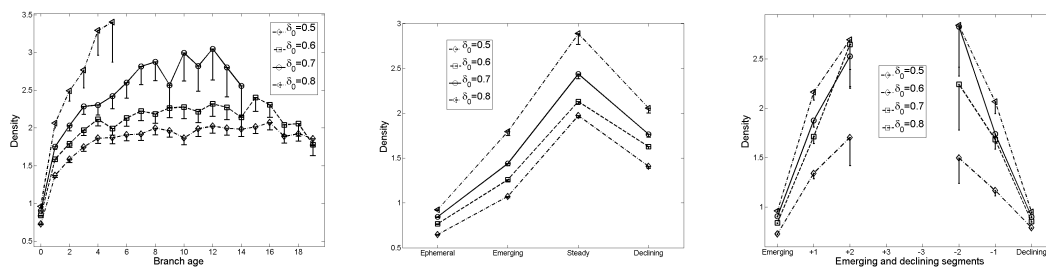


FIGURE 19 – Évolution de la densité d'un champ a) le long des branches phylométriques, b) en fonction de la typologie du nœud et c) sur les segment émergents et déclinants. Les différentes de valeurs de  $\delta$  correspondent à différents choix de reconstruction quant au seuil de l'appariement inter-temporel. Les régularités mises en évidence sont stable même lorsque le nombre de lien dans la phylométrie change d'un ordre de grandeur du fait de la variation de ce seuil.

### 3.2.3 Une méthodologie transversale

Ces travaux ont connu un certain succès (plus de 10.000 vues à ce jour sur PLoS One) et j'ai eu l'occasion d'en présenter les résultats dans plusieurs conférences et ateliers internationaux. Une phylométrie des sciences a par ailleurs été sélectionnée en 2013 pour faire partie de l'exposition *Places ans Spaces* coordonnée par l'Indiana University. Celle-ci rassemble chaque année depuis 10 ans les dix meilleures visualisations de connaissances [[Chavalarias et Cointet \(2013b\)](#)]. Cette approche est également référencée dans l'ouvrage *Atlas of Knowledge* [Börner \(2015\)](#).

Mes travaux ont été les premiers (en 2009) à proposer la reconstruction des dynamiques de connaissances à grande échelle dans le cadre d'une analyse par co-words (ce qui est aussi vrai, à ma connaissance pour les autres approches). Je les ai appliqués avec succès à plusieurs type de bases documentaires démontrant la possibilité d'étendre cette approche à divers contextes de production distribuée de connaissance :

- *Application à la blogosphère en 2010*, [Chavalarias et collab. \(2010\)](#) Collaboration avec le *Ministère de l'enseignement Supérieur et de la Recherche* (cf. fig. 20).

Ont été reconstruits sur 6 mois les débats dans le domaine de la santé et de l'environnement (sur une base de plus de 100.000 billets de blogs). Cette reconstruction est accessible via une plate-forme ([tina.iscpif.fr/veille/mesr](http://tina.iscpif.fr/veille/mesr) ; login : guest, mot de passe : iscpif).

- Application à la presse en 2011, *Chavalarias et collab. (2011a)* Collaboration avec l'initiative des Nations Unies Global Pulse pour la cartographie dynamique de la question de la sécurité alimentaire entre 2000 et 2011 (cf. fig. 21). Site web : <http://pulseweb.cortext.net>
- Application à l'évolution des thématiques des technologies futures et émergentes en 2013, *Chavalarias et Cointet (2013b)* Visualisation sélectionnée parmi les 10 meilleurs cartes de science de l'année dans l'exposition *Places and Spaces* coordonnée par l'Indiana University<sup>11</sup> et les travaux sur la phylométrie et la sécurité alimentaire sont référencée dans l'ouvrage "*Atlas of Knowledge*" Börner (2015).

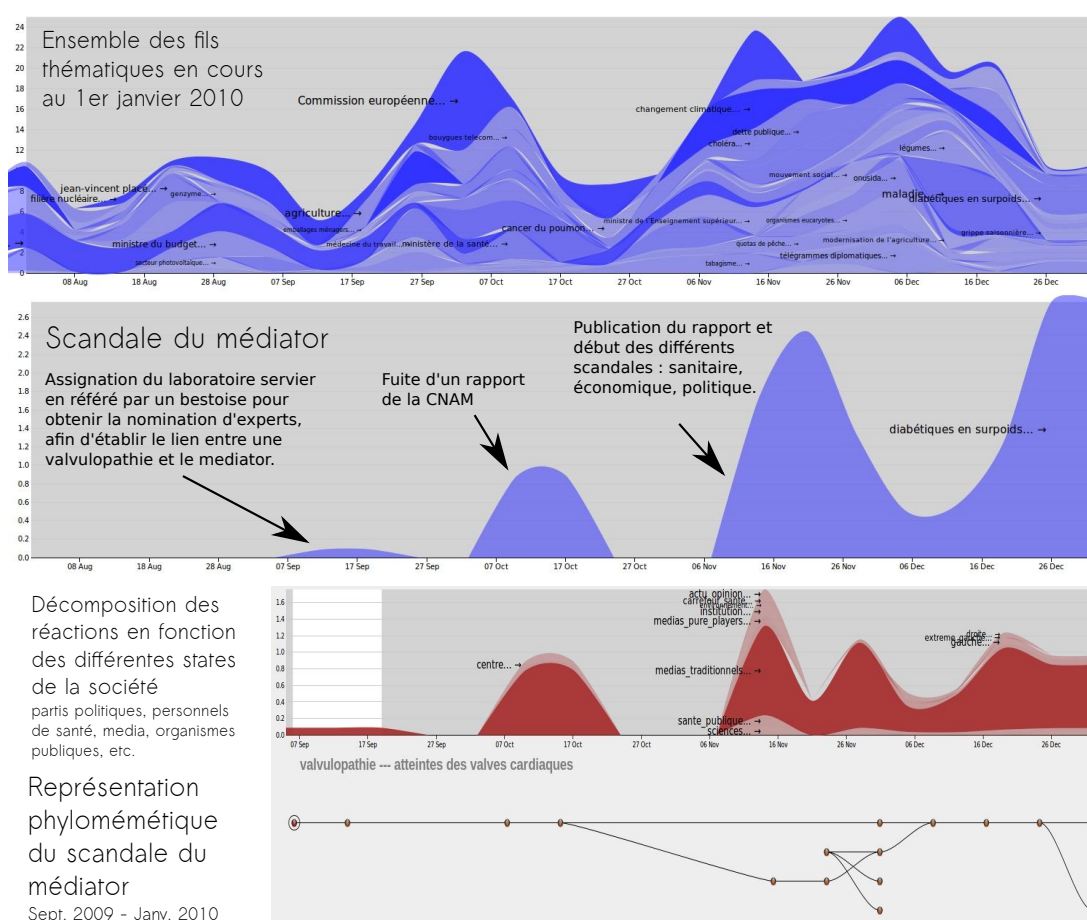


FIGURE 20 — Reconstruction des débats dans la blogosphère sur la période Août 2009 - Janvier 2010. Exemple du scandale du médiateur. *Chavalarias et collab. (2010)* Site : <http://tina.iscpif.fr/veille/mesr> ; login : guest, mot de passe : iscpif

11. Site web : [http://scimaps.org/mapdetail/science\\_phylomemy\\_159](http://scimaps.org/mapdetail/science_phylomemy_159)

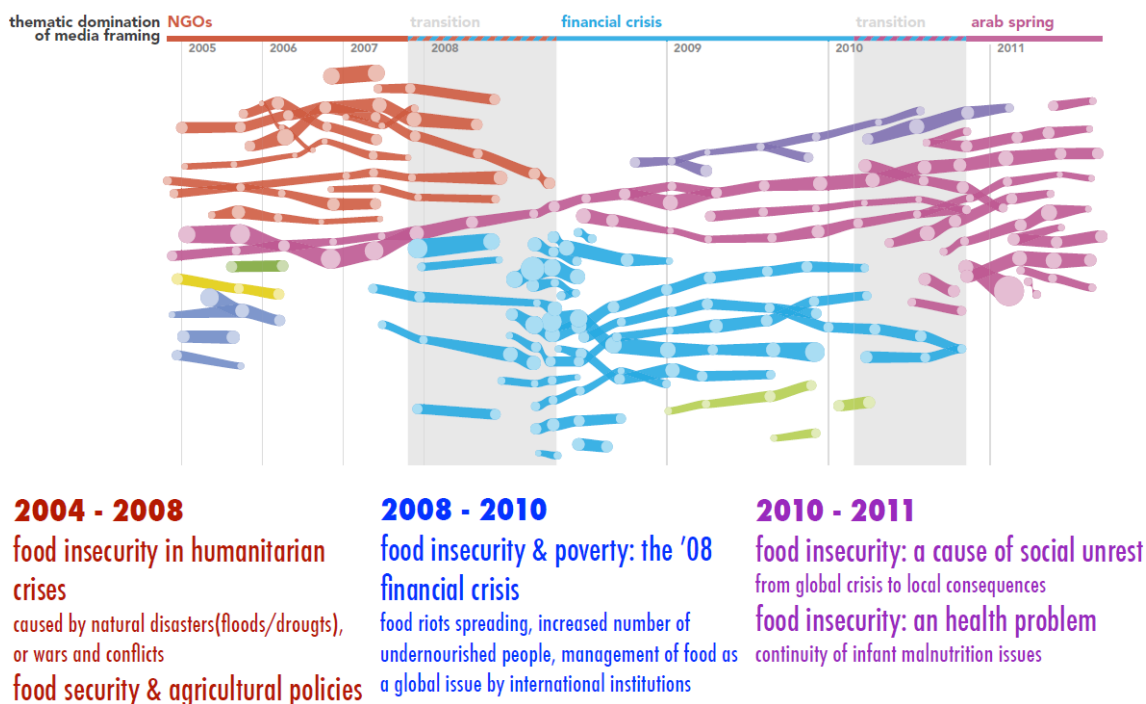


FIGURE 21 – Reconstruction du traitement du thème de la sécurité alimentaire par les médias entre 2004 et 2011. L'étude a mis en évidence les changements d'angle d'analyse suite à la crise financière de 2008 et au printemps arabe de 2011 (identifiés à travers les couleurs des branches phylogénétiques.) La thème de la sous-nutrition infantile est quant à lui présent sur toute la période. Chavalarias et collab. (2011a), Site web : <http://pulseweb.veilledynamique.com>

### 3.2.4 Vers une analyse morphologique de l'évolution des sciences

L'étude des phylomémies ouvre la voie à une étude quantitative de la science qui ne s'appuierait plus uniquement sur de simples nombres et leur évolution (nombre de publications, de citations, *h-facteurs*, etc.) mais également sur des éléments de *morphologie* des branches de sciences. Je propose de développer pour la scientométrie ce que la topologie a été aux mathématiques.

La figure 22 permet d'illustrer ce point de vue. Dans le cadre du projet européen TINA (2009-2011), j'ai eu accès à l'ensemble des projets soumis au programme interdisciplinaire *Future and Emerging Technologies* (FET Open) de la Commission Européenne (FP7). J'ai ainsi recueilli plus de 5000 mots-clés donnés par les auteurs des projets soumis à ce programme. J'ai effectué une reconstruction de phylométrie à partir de ces termes sur l'ensemble du Web of Science entre 1990 et 2010. L'un des domaines les mieux couverts par les projets de cet appel était le champ des *ordinateurs quantiques*, dont l'une des principales sous-branches est représentée sur la figure 22.

Ce champ peut-être délimité par des termes tels que *quantum computing*, *quantum computers*, *quantum processing* ou *quantum algorithms*. Il est particulièrement intéressant pour plusieurs raisons. 1) Il est extrêmement bien défini : il correspond à une branche exotique de l'informatique s'appuyant sur une nouvelle physique et possédant un vocabulaire très spécifique. 2) C'est un domaine où la théorie et l'expérience doivent aller de paire, chacune de ces activités étant bien représentée dans les publications. 3) C'est un domaine à fort enjeu sociétal et stratégique. Pour certaines tâches, les ordinateurs quantiques sont théoriquement des milliards de fois plus rapides que les ordinateurs classiques,

ce qui leur permettraient notamment de “casser” tous les protocoles de cryptographie utilisés actuellement. Les premières nations qui produiront un ordinateur quantique auront un avantage technologique majeur sur les autres, avec un accès à toutes les données privées ou confidentielles, raison pour laquelle l'armée américaine s'est impliquée très tôt dans ce domaine.

Le premier réflexe du scientomètre pour étudier l'évolution d'un champ scientifique consiste à en collecter le nombre de publications. Nous pouvons définir ici ce champ par la requête “(“*quantum computing*” OR “*quantum computer*” OR “*quantum processing*” OR “*quantum algorithm*”)”. Dans le Web of Science, nous observons que ce champ de recherche se constitue à partir de 1991, reste confidentiel jusqu'à 1995, émerge entre 1996 et 2002 et témoigne d'une activité soutenue depuis (cf. histogramme de la figure 22).

La reconstruction phylomématique donne un éclairage très différent sur cette évolution. La reconstruction présentée ici n'est que partielle, elle a été construite à partir d'une liste de termes qui proviennent exclusivement des projets *FET Open* sans intervention additionnelle et elle ne visait pas à être représentative du champ. Cette phylométrie met cependant en évidence, dès le premier coup d'œil, plusieurs phénomènes qui ne transparaissent pas dans l'analyse du rythme des publications.

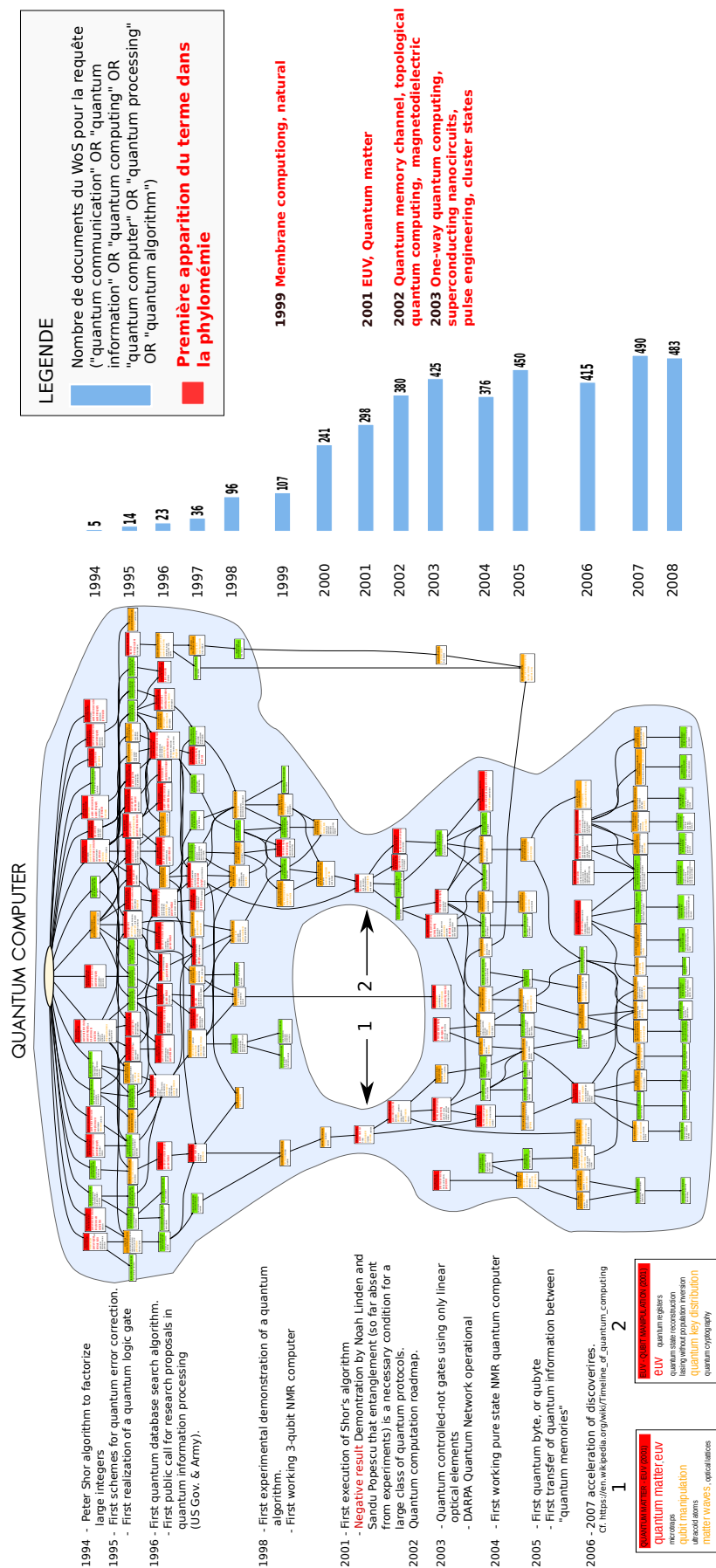
Comme on peut le voir sur la figure 22, il y a eu deux vagues d'innovations centrées sur les périodes 1992-1997 et 2001-2006 avec un “trou d'air” entre 1999 et 2001. La première vague d'innovation est très féconde en nouveaux concepts. Ceux-ci tiennent autant de la théorie (par ex. *spin cohérence* ou *quantum algorithms*) que des technologies de l'expérimentation (par exemple *optical lattices*, *semiconductors microlasers*). La deuxième vague d'innovations semble dominée par les technologies de l'expérimentation : *EUV (extreme ultra violet)*, *magnetodielectric*, *superconducting nanocircuits*, *One-way quantum computing*. Ceci laisse penser que le champ a rencontré un verrou technologique vers la fin des années 90, qui a été levée au début des années 2000, permettant le redémarrage du champ. Cette description de l'histoire du champ ne peut être inférée de la simple analyse de la productivité du champ en terme de nombre de publications, qui au contraire, donne l'impression d'un développement continu. Cette description est pourtant cohérente avec l'histoire de la discipline (cf. encart). Il y a eu effectivement en 1999 et 2001 deux résultats “négatifs” montrant à la fois que l'intrication quantique était une propriété nécessaire pour tout ordinateur qui aurait la prétention de supplanter les ordinateurs classiques, et que cette propriété était absente des dispositifs quantiques construits jusqu'alors (sur la technologie dite “NMR”). Ces limites ont été surmontées en quelques années avec l'invention du *one-way quantum computing*.

Cet épisode a néanmoins marqué la morphologie de cette branche, et la question se pose de savoir si la scission en deux branches qui forme un “donut” sur la période 1997-2004 est fortuite ou bien recèle une information sur la structure du champ à l'époque, les verrous technologiques qu'il devait surmonter, les pistes explorées pour les lever (ce qui semble être le cas au regard de la composition des champs). Nous ne pourrions répondre à ce type de question qu'en développant des analyses multi-échelles des réseaux dynamiques jouant sur la granularité de reconstruction de la phylométrie (notamment en faisant varier le seuil pour l'appariement inter-temporel). En effet, la structure en donut a toutes les chances d'être sensible à la granularité.

Cette première analyse, qui mériterait d'être beaucoup plus poussée, notamment en partenariat avec des chercheurs du domaine, met bien en évidence la richesse des perspectives qu'offre l'approche par reconstruction de phylométries.

Une autre analyse historique sur une autre branche de cette même phylométrie *FET Open* est présentée dans Chavalarias et Cointet (2013b) (domaine des neuroprothèses).

**FIGURE 22 – Phylométrie partielle du champ “Quantum computers”, 1991-2010.** Les champs (ensemble de termes labellisant une question de recherche) sont calculés sur des périodes de 5 années glissantes. Deux exemples sont donnés en 1 et 2 (en format numérique, zoomez sur n'importe quel champ pour voir sa composition). L'année indiquée dans l'entête est l'année médiane sur laquelle a été calculé un champ. Les entêtes rouges indiquent la présence de termes “émergents” [Chavalarias et Cointet (2013a)] (première apparition dans la phylométrie), celles oranges la présence de termes nouveaux par rapport aux antécédents directs du champ, mais présents dans d'autres branches de la phylométrie ou à des périodes anciennes ( $\geq 2$  ans, appelé *recombinaisons*), les entêtes vertes indiquent que tous les termes mentionnés étaient déjà présents dans les champs qui lui ont donné naissance. L'histogramme présente le nombre de publication dans le domaine des ordinateurs quantiques défini par la requête “*quantum communication*” OR “*quantum information*” OR “*quantum computing*” OR “*quantum computer*” OR “*quantum processing*” OR “*quantum algorithm*”). Les barres de l'histogramme sont alignées avec les périodes de la phylométrie. L'historique à gauche récapitule les principaux événements du domaine (voir aussi l'encart) mis en correspondance avec le déroulement de la phylométrie.



Elle met en évidence un autre motif remarquable dans l'évolution des sciences, à savoir l'hybridation entre deux champs scientifiques : le domaine des prothèses et de la réhabilitation au sens classique, et le domaine des sciences cognitives et des interfaces hommes-machines. Ces deux domaines ont convergé à la fin des années 1990/début des années 2000 pour constituer le champ des neuroprothèses, orthèses actives et interfaces cerveau-machine.



## Timeline of quantum computing 1990-2005

Source : *Wikipedia, the free encyclopedia*

[https://en.wikipedia.org/wiki/Timeline\\_of\\_quantum\\_computing](https://en.wikipedia.org/wiki/Timeline_of_quantum_computing)

Certains passages ont été écourtés pour des raisons éditoriales.

**1991**

- Invention of entanglement based secure communication.

**1993**

- Invention of an oracle problem for which a quantum computer would be exponentially faster than a conventional computer.

**1994**

- Peter Shor discovers an important algorithm to factor large integers quickly that could theoretically break many of the cryptosystems in use today. Its invention sparked a tremendous interest in quantum computers.
- First United States Government workshop on quantum computing.
- Proposition of an experimental realization of the controlled-NOT gate with trapped ions.

**1995**

- First United States Department of Defense workshop on quantum computing and quantum cryptography.
- Proposition of the first schemes for quantum error correction.
- Experimental realization of the first quantum logic gate with trapped ions.

**1996**

- Invention of the quantum database search algorithm.
- The United States Government and the National Security Agency, issues the first public call for research proposals in quantum information processing.
- Proposition a list of minimal requirements for creating a quantum computer.

**1997**

- First gates for quantum computers based on bulk spin resonance, or thermal ensembles. The technology is based on a nuclear magnetic resonance (NMR) machine.
- The principles of topological quantum computation are described as a method for combating decoherence.
- Loss-DiVincenzo quantum computer proposed,

**1998**

- First experimental demonstration of a quantum algorithm. A working 2-qubit NMR quantum computer used to solve Deutsch's problem was demonstrated.
- First working 3-qubit NMR computer.
- First execution of Grover's algorithm on an NMR computer.

**1999**

- Samuel L. Braunstein et al. showed that there was no mixed state quantum entanglement in any bulk NMR experiment. Pure state quantum entanglement is necessary for any quantum computational speedup, and thus this gave evidence that NMR computers would not yield benefit over classical computer. It was still an open question as to whether mixed state entanglement is necessary for quantum computational speedup.

**2000**

- First working 5-qubit NMR computer demonstrated.
- First execution of order finding (part of Shor's algorithm).
- First working 7-qubit NMR computer demonstrated.

**2001**

- First execution of Shor's algorithm. The number 15 was factored using 1018 identical molecules, each containing seven active nuclear spins.
- Noah Linden and Sandu Popescu proved that the presence of entanglement is a necessary condition for a large class of quantum protocols. This, coupled with Brauenstein's result (see 1999 above), called the validity of NMR quantum computation into question.
- Demonstration that optical quantum computing is possible with single photon sources, linear optical elements, and single photon detectors, launching the field of linear optical quantum computing.

**2002**

- The Quantum Information Science and Technology Roadmapping Project, involving some of the main participants in the field, laid out the Quantum computation roadmap.

**2003**

- Demonstration of quantum controlled-not gates using only linear optical elements.
- DARPA Quantum Network becomes fully operational.

**2004**

- First working pure state NMR quantum computer (based on parahydrogen) demonstrated.
- First five-photon entanglement demonstrated by Jian-Wei Pan's group, the minimal number of qubits required for universal quantum error correction.

**2005**

- Demonstration of quantum entanglement of multiple characteristics, potentially allowing multiple qubits per particle.
- Capacitance of a Josephson junction measured for the first time. The methods could be used to measure the state of quantum bits in a quantum computer without disturbing the state.
- First quantum byte, or qubyte, is announced to have been created.
- First success in transferring quantum information between "quantum memories" – from atoms to photons and back again.

### 3.3 Qualité des théories et évolution des normes de publication

Plusieurs observables macroscopiques permettent de quantifier la dynamique des sciences et le rythme des découvertes, les plus connus étant le rythme des publications, le temps moyen de citation, taille des communautés, etc. Le croisement, à l'échelle de l'histoire d'une communauté scientifique, de ces variables avec des marqueurs du rythme des découvertes (comme la publication d'une relation gène-protéine) a mené à des résultats très originaux comme par exemple He et Zhang (2009). Cependant, lorsque ces marqueurs sont moins évidents et moins normés, la corrélation entre ces observables et la qualité de la recherche est difficile à établir du fait même que la science comporte une dimension fondamentalement sociale : effets de mode, biais de publication, effets des agences de financement, etc.

La modélisation de l'évolution des sciences dans le cadre du jeu Nobel a permis d'émettre des conjectures sur la dépendance entre la rigueur des publications scientifiques et d'autres variables qualifiant l'environnement du chercheur : incitations à la publication, difficulté du domaine, taille de la communauté, etc. Je me suis alors naturellement intéressé à la question de l'identification de marqueurs de la qualité des publications et de leur évolution à partir d'une analyse empirique.

Dans de premiers travaux, j'ai mené une analyse systématique de l'ensemble des biais répertoriés dans la littérature biomédicale afin d'en faire l'histoire, mais aussi de mettre en évidence les associations entre biais et les biais potentiels associés à chaque thème de recherche [Chavalarias et Ioannidis (2010)]. Cela a donné lieu notamment à une base de données permettant de trouver les biais potentiels associés à un type de recherche.

De larges domaines des sciences utilisent de manière normée des indicateurs de fiabilité des résultats. C'est le cas en particulier de la majorité des recherches biomédicales qui utilisent les *P-valeurs* pour donner un indicateur de confiance dans leurs découvertes. Ce domaine a également été l'un des premiers à s'équiper d'archives numériques (MedLine) qui sont intégralement téléchargeables pour des besoins de text-mining. Dans Chavalarias et collab. (2016), j'ai mesuré de ce critère de qualité dans la littérature biomédicale à partir de l'extraction des valeurs de *P-valeurs* mentionnées dans les articles (résumés et textes pleins) et analysé sur 25 ans son évolution dans différents sous-domaines de la science biomédicale.

J'ai analysé l'ensemble de la base Medline (plus de 16 millions de résumés entre 1990 et 2015), ainsi que l'ensemble des archives plein texte de PMC (plus de 800.000 articles). C'est une des rares analyses de la littérature à cette échelle sur du texte plein. Les algorithmes que j'ai développés pour identifier les *P-valeurs* ont une précision estimée à 100% et moins de 2% de faux négatifs (*recall*=98%, sur 1000 articles vérifiés à la main). La base de données que j'ai constituée sur l'évolution des *P-valeurs* dans la littérature a un très fort potentiel scientifique pour de futurs travaux<sup>12</sup>. Elle peut par ailleurs être mise à jours très facilement car je me suis attaché à produire un workflow générique.

Avec mes collaborateurs, cette analyse nous a permis de démontrer l'existence de biais systématiques dans la présentation des résultats, l'objectif étant probablement de les rendre plus acceptables. Ces biais sont par ailleurs spécifiques, par leur importance et leur

12. Données téléchargeable sur <http://dx.doi.org/10.7910/DVN/6FMTT3>

évolution, en fonction des types de publications (meta-analyse, essais contrôlés randomisés, essais cliniques, revues) et pour différents niveau d'excellence de publications (core clinical journals et tout Pubmed). Enfin, nous avons montré qu'il y a un recours croissant à des P-valeurs non statistiquement significatives. Ces résultats seront à rapprocher des modèles d'évolution des sciences faisant des prédictions sur la relation entre qualité des publications, les normes scientifiques en vigueur et politiques scientifiques (voir section 3.1). Par ailleurs, j'ai l'intention de croiser cette approche avec la reconstruction de phylogénies afin de comprendre comment les normes d'utilisation de P-valeurs évoluent au cours de la constitution d'un champ scientifique.

Cette recherche est la première analyse de ce type sur les pratiques de justification statistique des résultats de recherche par une méthode qui est à la fois très utilisée par les chercheurs et très importante dans les processus d'évaluation des articles pour publication. Cette recherche est également un exemple d'analyse de l'évolution d'une norme de publication en science et une mesure à l'échelle de communautés entières, de la rigueur de la démarche scientifique et de son évolution.

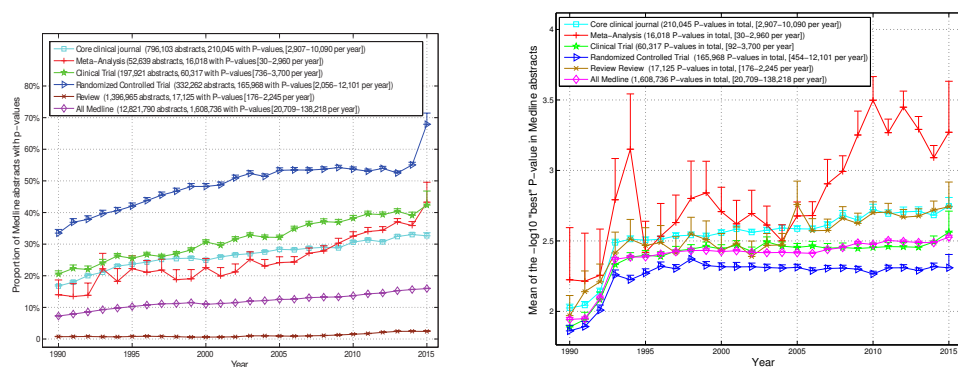


FIGURE 23 – A droite : Évolution des proportions d'articles faisant mention de P-valeurs dans leur résumé. A gauche : Évolution des moyennes des meilleures P-valeurs (minimale) mentionnées dans les résumés de MedLine par catégorie de publication. Statistiques réalisées sur 13M de publications. Les barres d'erreur sont les intervalles de confiance à 95%. Source : Chavalarias et collab. (2016)

### 3.4 Le Tweetoscope climatique : explorer les dynamiques culturelles autour d'un enjeu de société

La sphère académique est l'une des arènes de production culturelle, comment s'articule-t-elle avec la sphère citoyenne, notamment sur les questions qui leur sont communes ?

J'ai entamé en 2015, avec mon équipe de l'ISC-PIF, un travail sur ce thème autour du projet de *Tweetoscope Climatique* [Chavalarias et collab. (2015)], dispositif que nous avons créé à l'occasion de la COP21 et de l'exposition *Le Climat à 360°* (octobre 2015 - mars 2016, Cité des Sciences et de l'Industrie). Il est également disponible en tant que visualisation interactive sur <http://tweetoscope.iscpif.fr>

Côté science, nous avons analysé 58M de publications du Web of Science, dont 392.000 portaient sur le changement climatique pour produire une carte de "état de l'art" sur les trois dernières années des recherches sur le sujet (fig. 24).

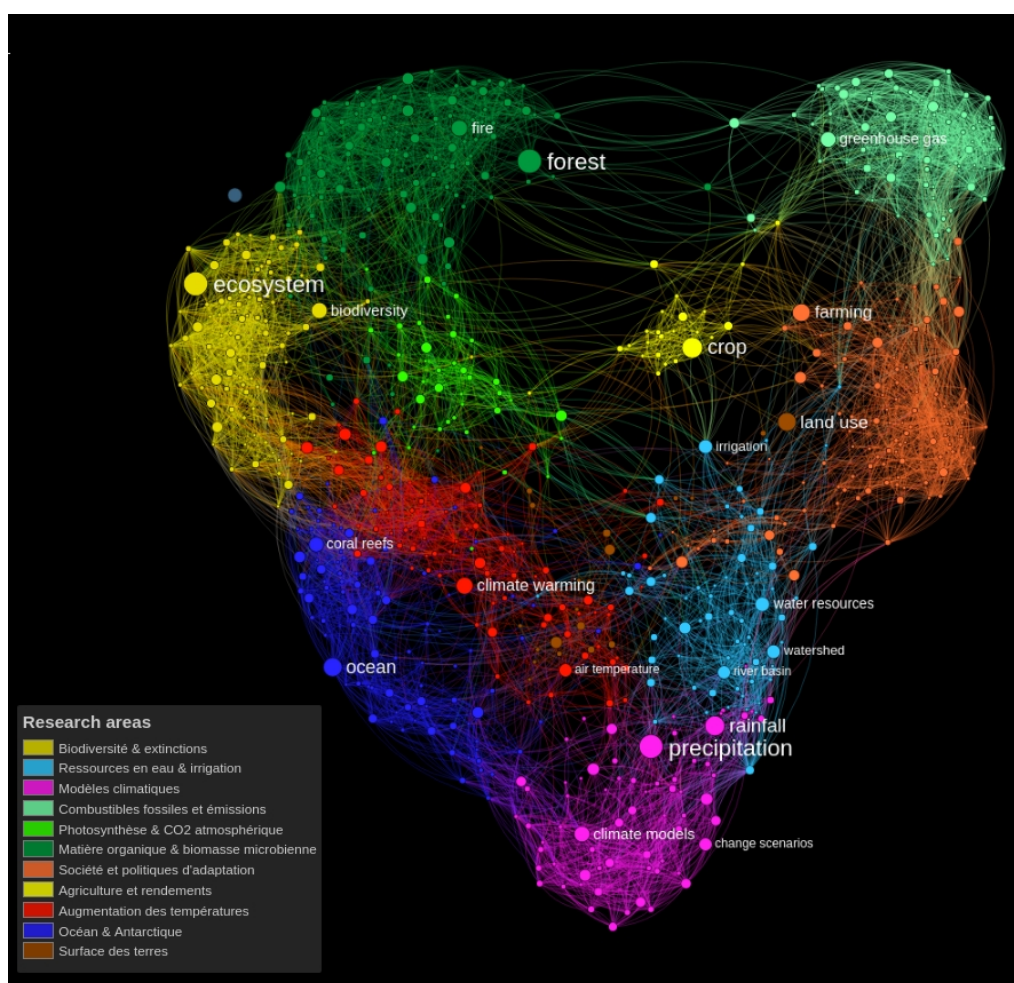


FIGURE 24 – Vue thématique du Tweetoscope qui permet d'explorer la littérature scientifique. Les termes spécifiques aux recherches sur le changement climatique sont représentés avec une taille proportionnelle au nombre de publications les mentionnant. Des liens les relient lorsqu'ils sont mis fréquemment en relation par les chercheurs. Une couleur est attribuée à chaque grand thème de recherche : gaz à effet de serre, ressources en eau et irrigation, augmentation des températures... <http://tweetoscope.iscpif.fr>

Côté citoyens, demeure la question de la connaissance de ces thèmes et de leur perception. Pour avoir des éléments de réponse, nous avons utilisé la plate-forme de micro-

blogging Twitter pour capter et analyser depuis mars 2015 un échantillon de ressources du web. A partir d'une centaine de mots-clé liés au changement climatique (changement climatique, global warming, etc.), nous avons collectés des tweets pertinents au rythme de quelques centaines de milliers par semaine. Ce corpus, mis à jour en temps réel, comporte actuellement plus de 20M de tweets.

Les tweets sont peu informatifs du fait du faible nombre de caractères qu'ils comportent et des abréviations utilisées. Par contre, ils mentionnent fréquemment des liens vers des pages web (articles de presse, blogs, articles scientifiques, sites institutionnels, etc.) qui correspondent à des ressources pertinentes pour les auteurs de ces tweets. Nous avons donc moissonné le contenu de ces pages web, constituant ainsi un corpus de 2,8M de documents (mars 2015-novembre 2015), dont 2,3M se sont révélés porter sur le changement climatique (i.e. contenant au moins un terme parmi une centaine de termes, français ou anglais, associés au changement climatique). Ce sont les contenus de ces pages web qui nous ont intéressés comme proxy de ce dont parlent les membres de ce réseau social (qui compte plus de 300 millions d'utilisateurs actifs) et plus généralement le web. Les 2500 termes provenant de la littérature académique sur le changement climatique ont été indexés dans ces contenus, ce qui nous a permis de les mettre en correspondance avec des termes de la carte.

Pour estimer l'importance relative accordée par le web à ces thématiques en comparaison avec ce qui se passe en sciences, j'ai considéré comme indicateur le ratio des proportions de documents dans lequel un terme figure côté sciences et côté web. Ce ratio peut être calculé à différentes périodes pour étudier sa dynamique.

Les premiers résultats montrent que la couverture du Web sur un thème comme changement climatique est très bonne, même pour des sujets scientifiques très techniques (de fait, il y a également des scientifiques qui tweetent sur ce sujet). Mais le fait le plus remarquable est la concentration de l'attention du web sur deux zones thématiques (cf. fig. 25) : la question de l'énergie, des émissions de gaz à effet de serre et de leur réglementation d'une part, les impacts sur les populations humaines d'autre part. En dehors de cette polarisation relativement stable, on observe des « vagues d'attention » relativement furtives dans ce paysage sémantique, vagues qui correspondent souvent à une actualité. Celles-ci peuvent être visualisés dans le Tweetoscope en jouant le « film » de l'évolution de l'attention semaine après semaine depuis mars 2015.

Ces différences dans les modes de l'attention de différentes strates de la société - communautés scientifiques, citoyens, politiques, sphère économique - posent la question de l'articulation des différentes composantes de la société qui, sur un même sujet tel que le changement climatique, ont chacune leurs priorités, leur agenda, leurs centres d'intérêts et leurs biais culturels propres. Ces strates ont également chacune leur propres temporalités. Si la position de la communauté scientifique évolue sur des unités de temps qui se comptent en mois (le temps de production d'un article), l'opinion publique est beaucoup plus versatile.

Cet axe de recherche n'en est qu'à ses débuts, mais l'analyse détaillée de l'évolution de l'attention du Web permettra de voir de quelle manière les différents acteurs du Web traitent les questions du changement climatique et comment ils se positionnent par rapport aux productions scientifiques. Nous avons ici un exemple typique de système dynamique lent/rapide. L'étude des influences entre les différentes parties de ce système (science, institutions, media, opinion publique) sera très instructive pour comprendre la manière dont la société dans son ensemble réagit face à un enjeu majeur comme le changement climatique.

Le Tweetoscope intègre aussi une dimension participative et réflexive, il s'enrichit au

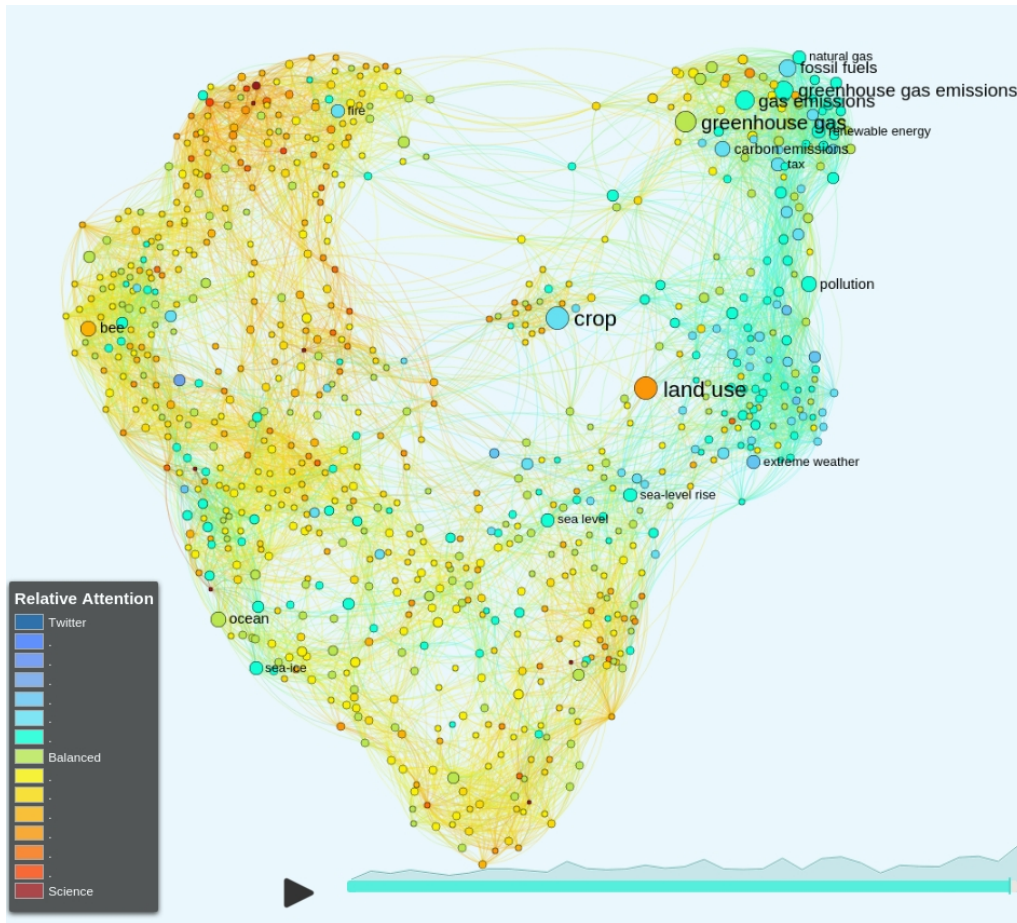


FIGURE 25 – Comparaison entre l'attention dans la sphère académique et l'attention sur le Web. La taille d'un terme est fonction de la proportion de documents qui lui est associée, tous supports confondus : WoS et Web. La couleur indique quant à elle les communautés qui s'intéressent majoritairement à ce terme : bleu pour le web (media, institutions et opinion publique), orange/rouge pour la communauté scientifiques, vert lorsqu'une égale attention est accordée par l'ensemble de ces communautés.

cours du temps avec l'arrivée de nouveaux corpus (scientifiques ou issus du web), mais aussi avec la prise en compte de suggestions de scientifiques et de citoyens qui peuvent proposer des ajouts ou des retraites de termes de la carte ou souligner l'importance ou la pertinence de certaines ressources. En tant que dispositif évolutif, il permet donc la construction d'une représentation collective des questions et enjeux liés au changement climatique et des réponses qui peuvent y être apportées. Cette représentation collective et partagée est à mon sens indispensable à l'établissement d'un dialogue constructif entre les différentes strates de la société.

Le Tweetoscope Climatique est donc un *macroscope* au sens de De Rosnay, dans toute sa dimension réflexive, un point de départ pour de nouvelles recherches sur les dynamiques culturelles et un support de réflexion pour l'action publique.

## 3.5 Enjeux de l'analyse quantitative de l'évolution des sciences et de l'évolution culturelle

### 3.5.1 Qualification multi-échelles de la morphologie des phylomémies

*Un des enjeux théoriques majeurs de ce programme de recherche est de qualifier les formes qui apparaissent à différentes échelles de la structure d'une phylométrie, évaluer leur stabilité et les relier aux motifs de la micro-dynamique conceptuelle des champs de recherche.*

En particulier, on étudiera la manière dont les concepts migrent d'une branche de phylométrie à l'autre, dans quels contextes se font les découvertes et les avancées théoriques, quelles sont les interactions entre types de renouvellement conceptuel et la morphologie des phylomémies (sujet déjà évoqué dans [Chavalarias et Cointet \(2013b\)](#)).

Pour ce faire, j'ai l'ambition de développer des reconstructions phylométriques à l'échelle de la science tout entière, en fusionnant des bases telles que le Web of Science, Medline, Istex, Repec, etc. Ceci permettra de procéder à une étude comparative des différentes branches de la science, et à une étude fine des transferts de concepts entre branches.

Comme je l'ai exposé dans les sections [3.2.1](#) et suivantes, la reconstruction des réseaux phylométriques, qui s'inscrit dans le cadre de l'analyse par co-termes, peut s'appliquer à une grande variété de media, puisqu'elle ne fait pas d'autres hypothèses sur les documents analysés que la présence d'un contenu textuel et d'une date de rédaction. En cela, les questions qui entrent dans le cadre de l'analyse des phylomémies peuvent être très différentes de celles posées traditionnellement en scientométrie, et sont associés à plusieurs autres enjeux importants.

### 3.5.2 Temporalités et circulation de l'attention sociale

Comme nous l'avons vu en [3](#), nous disposons actuellement d'importantes bases de données textuelles issues de divers types d'activités sociales. J'ai ainsi pu appliquer l'approche de reconstruction phylométrique à d'autres media comme la blogosphère [[Chavalarias et collab. \(2010\)](#)] ou la presse [[Chavalarias et collab. \(2011a\)](#)]. Dans le Tweetscope climatique, nous avons également comparé la dynamique des sciences et la dynamique des débats sur Twitter. Ces travaux ont mis en évidence des différences qualitatives des dynamiques de ces média. Elle a pour origine la nature de ce qui y est débattu autant que la nature des acteurs qui mènent ces débats. Pour produire un article scientifique, il faut constituer une expertise sur plusieurs mois, voire plusieurs années. Un champ de recherche met quelques années à émerger ou à se réorienter en raison de son inertie intrinsèque, conséquence du temps nécessaire de formation des chercheurs. À l'autre bout du spectre, sur les media sociaux, les citoyens expriment leurs opinions ou relaient l'information de manière beaucoup plus spontanée et versatile. Ils peuvent changer de sujet en quelques secondes ; une nouvelle controverse collective peut émerger en quelques heures.

De même que les sciences cognitives distinguent différents niveaux d'attention articulés avec la *mémoire à court terme* et la *mémoire à long terme*, on qualifiera empiriquement différents niveaux d'*attention sociale* liés aux processus de *cognition sociale* de nos sociétés, et développés par différentes catégories d'acteurs. Ces niveaux varient d'une attention court terme à une attention long terme, chaque type d'attention étant pris en charge par des

domaines d'activités sociales distincts (une même personne pouvant participer à plusieurs niveaux d'attention collective distincts).

Une société saine doit savoir articuler ces différents types d'attention sociale : les sciences ne peuvent ignorer les enjeux sociétaux sur le moyen ou long terme, les politiques ne peuvent ignorer les préoccupations de leurs citoyens, même sur le court terme, la société civile ne peut ignorer les enseignements de la science quant à sa viabilité et celle de son environnement.

*Un second enjeu de ce programme de recherche est de comprendre l'articulation des différents niveaux d'attention sociale à partir de reconstructions phylométriques croisées sur différents media : coexistence et interférence de différentes temporalités, circulation des sujets de débat, constitution des groupes d'acteurs autour de ces débats, organisation de la vie des individus à travers leurs différents rôles dans ces différentes strates de l'attention sociale, qualification des types de mémoire et d'oubli de ces strates, etc.*

Pour ce faire, il faudra non seulement continuer à développer les techniques de reconstruction de phylométries, mais aussi les hybrider avec d'autres approches comme l'analyse des citations, la *sentiment analysis* (encore balbutiante), le traitement automatique du langage ou les analyses de l'argumentation (comme par exemple *Prospero*, [Chateauraynaud (2003); Trabal (2004)]).

### 3.5.3 Réseaux phylométriques et dynamiques sociales

La plupart des documents à partir desquels sont reconstruits les phylométries comportent bien d'autres informations, souvent de type *meta-données*, qu'il est pertinent de re-analyser à l'aune de ces reconstructions et du *niveau d'observation* qu'elles créent. La méthode dite de "fond de carte" (*map overlay*) ouvre de nombreuses possibilités d'analyse.

Par exemple, si l'on dispose de mentions d'auteurs associées aux documents, on peut se poser la question de la "diffusion" de ces acteurs dans le paysage phylométrique. Cette idée est illustrée par la figure 26, qui présente une phylométrie dont les couleurs reflètent l'évolution des thèmes traités par les publications d'un chercheur au cours de sa carrière. Cette analyse peut être effectuée à différents niveaux d'agrégation : individuel, institution, pays, etc.

Cette approche permettra de définir des types de parcours (interdisciplinaire, théorique, empirique, explorateur, défricheur, consolideur, *mainstream*, etc.), dont on peut espérer qu'ils seront moins réducteurs dans la qualification des profils que les approches par *h-facteur* et autres index, qui ignorent la spécificité et le contexte de ce qui est produit. Une *approche qualitative des profils individuels* peut donc s'appuyer sur une *approche quantitative de la production collective*.

Au niveau collectif, on peut espérer mieux comprendre la manière dont les communautés scientifiques engendrent le paysage conceptuel de la science et s'adaptent à ses modifications. Par exemple, on pourra se demander comment la communauté des chercheurs en informatique et cybernétique des années 70 a "diffusée" dans un paysage qui s'est ramifié en champs aussi divers que l'intelligence artificielle, l'apprentissage machine, la fouille de données, les SVM ou les réseaux de neurones ; ou comment une communauté réagit à une innovation technologique comme les micro-array ou la microscopie confocale.



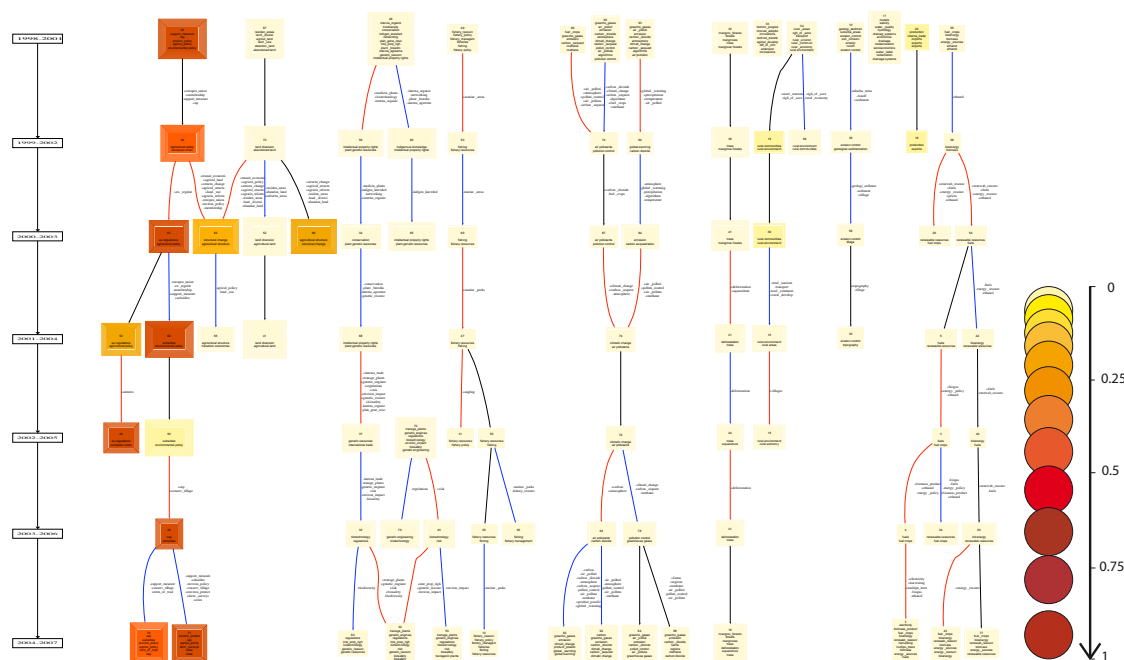


FIGURE 26 – Exemple d'utilisation d'une phylométrie pour analyser le parcours d'un chercheur. Les champs d'une phylométrie sur le développement durable sont colorisés en fonctions du nombre de publications de ce chercheur qui sont pertinentes pour ces champs. On peut ainsi voir l'évolution thématique du chercheur au cours du temps. Source : Cointet et Chavalarias *European Conference on Complex Systems 2008*, non publié.

Le niveau d'observation proposé par les phylométries sera pertinent si, par exemple, il permet de décrire ces processus d'évolution de communautés, d'innovation ou de découverte de manière plus synthétique, ou d'en accroître la prédictibilité, comme ont pu le faire Roth et Cointet (2010) sur les réseaux socio-sémantiques.

### 3.5.4 Réseaux phylométriques multi-partites

Mon expérience dans le domaine de la cartographie des sciences m'a convaincu que les phylométries constituent un *niveau d'observation* pertinent pour l'étude des dynamiques informationnelles au sein de nos sociétés. La question se pose cependant de savoir si c'est également un *niveau d'organisation*<sup>13</sup> pertinent pour les sciences et la société en général. Il ne s'agit pas de savoir si les phylométries ont une existence matérielle de quelque sorte, mais si elles nous permettent de synthétiser des processus, physiques ou immatériels, susceptibles d'interagir avec d'autres processus *macro*, comme la formation d'institutions, *meso* comme la formation de groupes de recherche ou de collaborations, ou *micro*, comme le cheminement intellectuel d'un chercheur ou l'évolution de sa carrière. Pour ce faire, l'enjeu est de contextualiser les réseaux phylométriques dans un ensemble plus large incluant les réseaux et les groupes sociaux, les institutions, etc. de manière à formaliser des approches multi-niveaux telle que celle développée par exemple par Lazega (2016) ou Lazega et Snijders (2016) à partir d'enquêtes sociologiques. On parlera alors de contraintes structurelles dans les *réseaux phylométriques multi-partites*<sup>14</sup>.

13. Au sens de la réflexion présentée en 1.1.2.

14. Qui sont également multi-niveaux et multi-échelles.

### 3.5.5 Réseaux phylométriques et modélisation de l'évolution des sciences

La motivation initiale qui m'a conduit vers l'analyse quantitative des sciences reste l'un des horizons de ce programme de recherche : amener les reconstructions de la dynamique des sciences à un niveau où elles pourront corroborer ou infirmer des théories de philosophie ou d'histoire des sciences, telles que celles formulées par Kuhn (1962), Popper (2002, 2006), Hull (1988) ou Danermark (2003) pour n'en citer que quelques-unes. Mon objectif est donc que mes futurs travaux autour de la modélisation des théories de Popper, et leurs éventuelles adaptations à d'autres théories de la connaissance, puissent être mis en vis-à-vis avec les résultats des reconstructions phylométriques qui en constitueront des tests empiriques.

## 3.6 Feuille de route technologique pour l'étude phylométrique des sciences

Plusieurs avancées technologiques sont nécessaires pour relever les défis que nous venons de mentionner. Les reconstructions de phylométries qui ont été effectuées jusqu'à ce jour portaient uniquement des branches de la science (embryologie, nouvelles technologies, etc.). A cette échelle, un chercheur peut vérifier "à la main" la cohérence du périmètre sémantique proposé par les algorithmes de text-mining (<10k termes) et un serveur peut effectuer les reconstructions en temps raisonnable sur quelques centaines de milliers de documents. Ces opérations deviennent hors de portée lorsque l'on passe à l'échelle sur l'ensemble de la science, ce qui implique des développements spécifiques.

### 3.6.1 Passage à l'échelle des algorithmes de reconstruction

Outre le fait que les algorithmes de reconstruction continueront à faire l'objet d'une recherche en tant que telle (qualité et variété des reconstructions), leurs performances en terme de rapidité devront nécessairement gagner plusieurs ordres de grandeurs pour traiter une proportion significative de la science (estimées à  $\simeq 100$  millions de documents et 20 millions d'expressions). J'ai lancé en 2016 une thèse sur ce sujet, en collaboration avec Bernd Amann et Hubert Naacke (LIP6, UPMC) ; et j'ai développé, grâce à la Région Île-de-France, des infrastructures de calcul appropriées. L'objectif est d'effectuer ce passage à l'échelle dans les trois ans.

### 3.6.2 Crowdsourcing et visualisation interactive

Les approches purement algorithmiques ne seront pas suffisantes quantitativement et qualitativement.

Quantitativement, parce que les algorithmes de text-mining, aussi performants soient-ils, ont toujours un taux d'erreur non négligeable sur la tâche qu'on leur assigne. Sans intervention humaine, le périmètre sémantique d'une phylométrie (liste des termes qui sont considérés pour la reconstruction) contiendra nécessairement des scories, expressions incomplètes ou peu informatives qui brouilleront les reconstructions.

Qualitativement, parce que de même que je ne crois pas à l'existence de LA carte

des sciences<sup>15</sup>, il y a une grande variété de reconstructions phylomémétiques<sup>16</sup> dont la pertinence est fonction de l'objet d'étude et de la question de recherche posée. À côté de structures robustes que l'on peut s'attendre à découvrir pour un large ensemble de catégories de reconstruction phylomémétique, il y aura une multitude de structures qui seront mises en évidence pour des catégories particulières, et dont on peut s'attendre à ce qu'elles intéressent particulièrement certains types d'acteurs. Pour que les phylomémies remplissent leur rôle en tant que supports pour la représentation et l'interprétation, il est nécessaire de déployer des mécanismes de co-construction qui impliquent les utilisateurs finaux.

Ce type d'approche est déjà à l'œuvre dans deux projets que je coordonne à l'ISC-PIF, le logiciel Gargantext [Delanoë et collab. (2015)] et le projet *Science en Poche*. Il devra néanmoins être développé substantiellement pour relever le défi du passage à l'échelle.

---

15. Cf. mes remarques section 3.2.1

16. Choix des types de proximités inter-termes, des types de clustering, de la granularité et surtout du lexique.

### 3.7 Résumé

Mes recherches dans le domaine de l'analyse de la dynamique des sciences ont contribué au développement de différents domaines :

- **Informatique** J'ai proposé de nouvelles méthodes de reconstruction des *réseaux dynamiques* qui sont aussi de nouvelles manières de faire du clustering dans les grands graphes sémantiques. Elle ouvrent de nouvelles possibilités dans le domaine de la *recherche d'information*. J'ai également développé de nouvelles approches dans le domaine de la *visualisation d'information*, qui permettent d'interagir avec les structures dynamiques de la science. Ces recherches ont de nombreuses applications dans le domaine de la veille technologique, les bibliothèques numériques ou de la gestion de la recherche (cf. section 6).
- **Sciences Humaines et Sociales** Mes travaux ont pour ambition d'amener les reconstructions phénoménologiques de l'évolution des sciences à un niveau où elles pourront servir de test, au sens de Popper, aux théories avancées en histoire et philosophie des sciences. D'un côté, ils traduisent en modèle des théories existantes en philosophie et histoire des sciences de manière à formuler des prédictions vérifiables ; de l'autre, ils reconstruisent l'évolution des sciences à partir des données dans le but de formuler des descriptions quantitatives et qualitatives de cette évolution. Ces recherches apportent un point de vue complémentaire à celui de l'histoire des sciences traditionnelle, qui est souvent celle des "Grands Hommes", en proposant une histoire des sciences quantitative écrite à partir de la multitude des contributions de la communauté scientifique.
- **Biologie et médecine** Mes travaux ont apporté des éclairages nouveaux sur les pratiques de l'utilisation des P-valeurs, ainsi que sur les types de biais associés aux différents domaines de recherche.
- **Interface science/citoyens** Les différentes plate-formes que je développe contribuent à faire le pont entre la communauté scientifique et les citoyens, là où un dialogue est nécessaire pour traiter des enjeux sociétaux (ex. le projet de Tweeposcope Climatique).

## PUBLICATIONS ET TRAVAUX ASSOCIÉS

- Chavalarias, D., Cointet, J.-P., 2008. Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study. *Scientometrics* 75, 37–50. doi :10.1007/s11192-007-1825-6
- Cointet, Jean-Philippe, Chavalarias, David, 2008. Multi-level Science mapping with asymmetric co-occurrence analysis : Methodology and case study. *Networks and Heterogeneous Media* 267–276.
- Funabashi, M., Chavalarias, D., Cointet, J.-P., 2009. Order-wise Correlation Dynamics in Text Data, in : Fortunato, S., Mangioni, G., Menezes, R., Nicosia, G. (Eds.), *Complex Networks : Results of the 2009 International Workshop on Complex Networks* (CompleNet 2009), Studies in Computational Intelligence. Springer, Berlin.
- Chavalarias, D., Ioannidis, J.P.A., 2010. Science mapping analysis characterizes 235 biases in biomedical research. *Journal of Clinical Epidemiology* 63, 1205–1215. doi :10.1016/j.jclinepi.2009.12.011
- Chavalarias, D., Cointet, J.-P., Roth, C., 2010. Étude portant sur un outil de veille dynamique sur l'attente sociétale en matière de recherche et d'innovation (Recherche et Développement). *Ministère de l'Enseignement Supérieur et de la Recherche*, Paris.
- Chavalarias, D., Jean-Philippe, C., T., C.L., Duong T. K, Mogoutov A., Villard L., Roth C., Savy T., 2011. Streams of Media Issues - Monitoring World Food Security, *UN Global Pulse White Paper*, United Nations.
- D. Chavalarias et J.-P. Cointet, « Phylomemetic patterns in science evolution—the rise and fall of scientific fields », *PloS one*, vol. 8, n2, p. e54847, 2013.
- Chavalarias, D., Castillo, S., Panahi, M., 2015. Le Tweetoscope Climatique. Une représentation collective des enjeux autour du climat. *Lettre de l'INSHS* 38, 39–42. <http://tweetoscope.iscpif.fr>
- Chavalarias, D., Wallach, J. D., Ho Ting Li, A., Ioannidis, J. P. A. (2016) Evolution of P-values in the biomedical literature. *En évaluation à JAMA*.

---

## 4 Modéliser l'évolution culturelle

---

La sociologie (néo-)structurale a depuis longtemps mis en évidence l'importance de la formation et du rôle de structures intermédiaires entre le niveau micro et le niveau macro pour la compréhension de nos sociétés. Ces structures ont été modélisées avec des formalismes comme le *block modelling* [Boorman et White (1976); Breiger (1974)] ou les *réseaux multi-niveaux* [Lazega et Snijders (2016)]. La question de la co-évolution des identités et des groupes sociaux est par ailleurs un débat sociologique majeur [Stryker et Burke (2000)] : les individus par leurs interactions forment des réseaux en choisissant d'après leurs préférences de créer, de défaire des liens ou de s'affilier à des groupes ; ces liens et ces affiliations, par les interactions qu'ils permettent, sont vecteurs de la modification des préférences, croyances et représentations des individus via des processus d'influence sociale.

L'opposition entre les approches holistes de la sociologie et l'individualisme méthodologique prôné par l'économie pourrait paraître caricaturale de nos jours, les frontières entre ces deux disciplines étant devenues relativement perméables. Néanmoins, mis à part des travaux tels que Horst et collab. (2006), les entreprises de modélisations tendent à adopter l'une ou l'autre des ces approches.

Mes travaux dans ce domaine ont pour objectifs de contribuer à la proposition d'un point de vue intermédiaire prenant en compte la co-construction des individus et du collectif, formalisant l'articulation entre comportements individuels et comportements collectifs, entre le niveau macro et le niveau micro, explicitant si besoin des structures meso. Il s'agit de modéliser le couplage entre l'évolution des structures sociales, la *formations des réseaux socio-économiques*, et des identités des agents pris au sens large (*formation des préférences*, des représentations, des traits culturels, etc.). Les couplage de ces deux types de processus forment conjointement ce que l'on peut appeler *cognition sociale* (cf. section 2.3).

Ces recherches s'inscrivent dans le cadre de la micro-économie cognitive et de la sociologie.

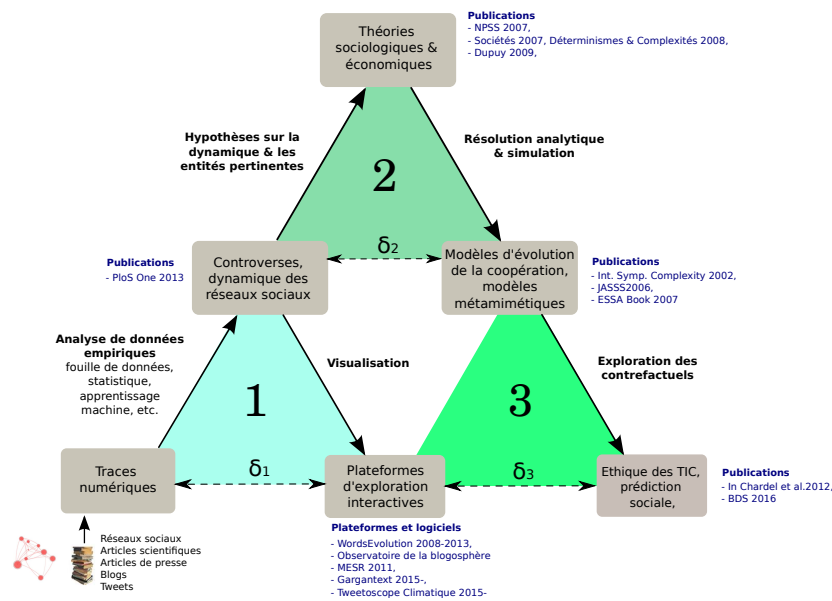


FIGURE 27 – Résumé de mes contributions dans l'étude de la cognition sociale

## 4.1 Modélisation de la formation endogène des réseaux

Ces recherches ont pour objectif d'étudier l'influence des réseaux sociaux et de leur topologie sur les processus socio-économiques. Je me suis placé dans le cadre de la modélisation des *dilemmes sociaux* et plus particulièrement du *dilemme du prisonnier* comme situation prototypique pour étudier ce phénomène.

Dans mes premiers travaux (2000-2002), j'ai modélisé une population d'agents en réseau en situation de dilemme du prisonnier dans le cadre de la *théorie des jeux évolutionnaires* [Weibull (1995)]. Cette approche permet d'étudier des phénomènes d'*émergence de coopération* ou de la confiance dans des configurations où il y a conflit entre intérêt collectif et intérêts individuels.

La littérature en économie expérimentale ou sociologique rapporte de nombreux exemples où une partie de la population adopte des comportements coopératifs même en présence de comportements non coopératifs. Cette observation n'était pas compatible avec les résultats de la théorie des jeux évolutionnaires qui faisait systématiquement l'hypothèse de *panmixie* : les agents se rencontrent au hasard. Cette hypothèse a été relaxée de différentes manières. Nowak et May (1993) et Nowak et collab. (1994) ont étudié des agents avec des positions fixes sur une grille torique et ont mis en évidence l'effet de la topologie sur les situations de dilemme social. Pour autant, l'émergence de la coopération restait confinée à des situations très particulières et les attracteurs hétérogènes étaient rares. Stanley et collab. (1995) ont quant à eux fait l'hypothèse de la possibilité de sélectionner les partenaires et de refuser des interactions, ce qui rend la coopération viable mais au prix d'une hypothèse très forte sur la possibilité de filtrer les interactions. Dans les deux cas précédents, la coopération est stabilisée par des effets de *cliquishness* du sous réseau des agents coopérateurs.

Dans Ahn et collab. (2002); Chavalarias (2002, 2006a) j'ai montré que l'on pouvait proposer une modélisation plus parcimonieuse que le modèle de Stanley et collab. (1995), tout en conservant le fait stylisé d'hétérogénéité des attracteurs : retrait de la possibilité de refuser une interaction ; robustesse des résultats vis à vis de perturbations sur le choix des

actions. J'ai également démontré qu'en présence de défecteurs inconditionnels, il fallait au minimum deux autres types différents d'agents pour que l'émergence de la coopération soit possible : coopérateurs inconditionnels et réciprocaturs (les autres approches s'intéressaient jusqu'alors au maintien de la coopération et non à son émergence, ce qui est beaucoup plus difficile à expliquer). J'ai mis en évidence les structures remarquables du mécanisme proposé, qui se différencient des structures identifiées dans les modèles précédents. Dans les modèles que j'ai étudiés la viabilité des agents coopérateurs est due à une activité sociale plus intense ou une *popularité* plus importante plutôt qu'à une intégration dans un sous-réseau à forte proportion d'agents coopérateurs (*cliquishness*). Ces deux types d'explications (*popularité* et *cliquishness*) s'appliquent à des types d'activités sociales différentes et nous disposons aujourd'hui d'un nombre suffisamment de jeux de données sur les réseaux sociaux pour les tester empiriquement les prédictions de ces diverses classes de modèles, et les associer avec les configurations sociales les plus pertinentes.

Mes publications ont été parmi premières à modéliser la formation endogènes de réseaux dans le cadre de la théorie de jeux évolutionnaires et à démontrer formellement l'importance de la topologie des réseaux dans la structuration des phénomènes économiques et sociaux. Cette approche est devenue populaire bien plus tard, notamment en micro-économie.

Je n'ai cependant pas poursuivi cette voie de modélisation (et de publication) car le cadre de la théorie des jeux évolutionnaires m'a paru inapproprié pour la modélisation des dilemmes sociaux des sociétés humaines. Je ne pense pas que les mécanismes de sélection naturelle ou économique tels qu'ils sont mis en œuvre dans ces approches soient les bons moteurs de l'évolution culturelle. Il est nécessaire de modéliser l'évolution des préférences, des valeurs, des identités des agents ; plutôt que celle de leurs "gènes" ou de leurs comportements. Cette insatisfaction m'a amené à proposer le formalisme des *jeux metamimétiques* (cf. 4.2) qui modélisent les processus de cognition sociale plutôt que les processus de sélection d'agents.

#### PUBLICATIONS ET TRAVAUX ASSOCIÉS

- Chavalarias D. (2006) Cooperation : popularity or cliquishness ?, *Proceedings of European Conference on Complex Systems ECCS06*, Oxford sept. 2006
- Chavalarias D. (2002) Emergence of cooperation and selection of interactions, *International Symposium On Complexity*, Shanghai, août 2002.
- Ahn T.K. , Chavalarias D. and Lazar A. (2002) Endogenous Network Formation and the Evolution of Preferences, In David Sachar, ed. *Proceedings of the Workshop on Agent Simulation : Applications, Models, and Tools*. Argonne, IL : Argonne National Laboratory.



## 4.2 La cognition sociale et la question de la différenciation

*Economists and to a lesser extent other social scientists have resisted addressing the complexity and endogeneity of human motivations, not because we think the behavioral simplicity of homo economicus is an adequate representation, but rather because we lack both adequate conceptual tools and empirical information on the process of preference formation.*

*Bowles (2001)*

### 4.2.1 Entre individualisme méthodologique et holisme

La description des processus (collectifs) de formation des *valeurs* que nous attribuons aux différents objets et situations que nous rencontrons est restée longtemps hors du champ de la modélisation du fait de l'absence de données, expérimentations et formalismes adéquats. Le terme *valeurs* doit être pris ici non dans son sens économique (équivalent aux prix) mais dans le sens plus large des *valeurs socio-culturelles*. Il s'agit de la question de la formation des motivations, désirs, préférences ou convictions éthiques et philosophiques, dont on peut questionner la représentation dans le cadre d'une modélisation (par des préférences, des fonctions d'utilité, des règles de comportement).

Deux conceptions opposées dominent la modélisation en sciences sociales selon que ces valeurs socio-culturelles sont pensées comme des propriétés intrinsèques aux objets ou comme le reflet de normes sociales qui transcendent les individus. Le moteur des dynamiques sociales est alors recherché soit au sein des individus, dans des motivations et préférences qui préexistent à leur participation à la société, soit au contraire au sein d'institutions et de faits sociaux qui contraignent voire déterminent le comportement des individus. On aura reconnu ici l'opposition entre l'individualisme méthodologique, au centre de la plupart des approches en économie, et l'approche holiste longtemps dominante en sociologie.

Mes recherches s'inscrivent dans une voie médiane qui, bien que minoritaire, est défendue dans diverses branches des sciences sociales : la sociologie tardienne [Tarde (1890)], la psychologie sociale [Stryker et Burke (2000)], l'anthropologie girardienne [Girard (1977)], l'économie des conventions [Orléan (2011)] ou l'économie des institutions [Dolfsma (2004)] pour n'en mentionner que quelques unes. Toutes ont en commun d'affirmer que les traits des agents (selon les disciplines : valeurs, motivations, désirs ou préférences, etc.) sont co-construits au cours de leurs interactions sociales, cette co-construction pouvant être influencée par des prédispositions individuelles ou des structures sociales pré-existantes. La voie de modélisation que j'ai suivie pour rendre compte de cette co-construction est la modélisation des processus d'apprentissage social (*social learning*), et plus particulièrement la modélisation d'une notion généralisée d'imitation [Chavalarias (2009)].

Notons que cette approche s'inscrit dans une conception non téléologique de la cognition, défendue par Tarde en son temps [Tarde (1890)] :

“La fonction cérébrale, l’esprit, se distingue des autres fonctions en ce qu’elle n’est pas une simple adaptation à une fin précise par un moyen précis, mais une adaptation à des *fins multiples et indéterminée* [...]” [mes italiques]

Conception que l’on retrouve également en économie, chez des penseurs comme [von Hayek et collab. \(1978\)](#) :

“What we call mind is not something that the individual is born with, as he is born with his brain, or something that the brain produces, but something that his genetic equipment (e.g. a brain of a certain size and structure) helps him to acquire, as he grows up, from his family and adult fellows by absorbing the results of a *tradition that is not genetically transmitted*.” [mes italiques]

#### 4.2.2 L’imitation, trait d’union entre l’individu et le collectif

À la fin du XIX<sup>ème</sup> siècle, les sciences humaines et sociales ont connu un engouement sans précédent pour l’imitation en tant que principe structurant du social. Alors que James Mark Baldwin [[Baldwin \(1897\)](#)] plaçait l’imitation au centre de sa théorie du développement, et notamment de l’émergence de la volition, Gabriel Tarde [[Tarde \(1890\)](#)] promettait aux lois générales qui régissent la répétition imitative un rôle analogue en sociologie à celui de l’hérédité en biologie.

Ces deux approches se rejoignent autour de deux convictions, d’une étonnante actualité, dont la portée épistémologique est de première importance pour les sciences sociales :

1. Il y a dans les systèmes sociaux co-détermination des individus et du collectif, celle-ci étant rendue possible par le jeu de l’imitation. Les systèmes sociaux engendrent ainsi de façon interne leur propre espace de significations qui est le produit, au niveau collectif, d’un processus dialectique intrinsèquement social de constitution du sens. Celui-ci commence dès l’émergence de la distinction entre soi et autrui chez l’enfant : « *Le sentiment du moi se développe par l’imitation des autres*, écrivait Baldwin, *et le sentiment de l’autre s’enrichit en proportion du sentiment du moi.* »<sup>17</sup> [Baldwin \(1894\)](#)
2. Du point précédent dérive l’idée d’une spécificité des dynamiques sociales par rapport à celles rencontrées dans les autres systèmes naturels. La raison principale de cette conviction, pour Tarde comme pour Baldwin, est que ces dynamiques sont par essence de nature mimétique. Ceci implique qu’une théorie du social doit se doter de ses propres outils conceptuels et formels. Il serait ainsi trompeur de chercher à donner aux sciences sociales un « air biologique » ou un « air mécanique » en s’appuyant exclusivement sur les outils formels et conceptuels d’autres disciplines.

Le point 1) annonce un programme de recherche qui se déploie entre individualisme méthodologique et holisme, préfigurant l’*individualisme méthodologique complexe* introduit par [Dupuy \(1992\)](#). La tension entre les deux approches à première vue antithétiques du social est résolue par un raffinement de la notion d’imitation, qui se déploie le long d’un continuum embrassant à la fois des formes favorisant les déterminations sociales et des formes favorisant l’originalité individuelle. Il y a notamment chez Tarde et Baldwin

17. « *My sense of myself grows by imitation of you, and my sense of yourself grows in terms of my sense of myself* »

la conviction que l'imitation délibérée est le premier signe distinctif de la volition et de l'action raisonnée. Ainsi [Baldwin \(1897\)](#) (p. 393) écrivait :

Lorsque la volition apparaît, l'enfant apprend à se connaître lui-même comme agent. Il prend conscience de ses pouvoirs d'acquisition et de résistance en même temps que de sa suggestibilité. Puis, au milieu du conflit de l'imitation (coutumière) et de la suggestion (novatrice) l'enfant arrive à prendre conscience de soi comme agent libre et capable d'initiative. L'enfant ne veut pas imiter, du moins ce qu'on lui suggère d'imiter, mais se choisit lui-même des modèles d'imitation. Il est vrai que par la suite il imite quand même : mais cette imitation diffère du tout au tout de l'imitation suggérée

En contrepoint de cette citation, Tarde soutenait à la même époque, dans ses *Lois de l'imitation* (op. cit., Ch IV, p. 71) :

« Plus les suggestions de l'exemple se multiplient et se diversifient autour de l'individu, plus l'intensité de chacune d'elles est faible, et plus il se détermine dans le choix à faire entre elles, par des références tirées de son propre caractère, d'une part, et, d'autre part, en vertu des lois logiques que nous exposerons ailleurs ».

Chez ces deux auteurs, la manière dont l'individu se démarque de son environnement social tient donc précisément à une modification de la forme de l'imitation. D'une forme d'influence sociale innée commune à tous, l'imitation prend peu à peu des formes propres à l'individu, reflétant, en quelque sorte, sa personnalité. La différenciation sociale apparaît alors comme un espace de variation libre créé à l'intérieur même du domaine réservé aux processus d'influence sociale.

Quant au point 2), il est important d'un point de vue heuristique, car il invite à sortir des cadres traditionnels de modélisation. L'intuition d'une spécificité du social est aujourd'hui confortée par les difficultés rencontrées par les entreprises de modélisation telles que l'économie « orthodoxe » ou la sociobiologie, à expliquer certains faits sociaux à partir de modèles inspirés de la physique ou de la biologie.

### 4.2.3 Désir mimétique et imitation rationnelle

Le programme de recherche lancé parallèlement par Tarde et Baldwin n'a pas trouvé de successeur immédiat. L'acharnement de Durkheim à enterrer au plus vite les idées de Tarde n'y est pas totalement étranger. Après avoir été quelque peu délaissé, le thème de l'imitation a cependant été repris par de grandes figures des sciences sociales : [Pia-  
get et collab. \(1945\)](#) en psychologie, [Girard \(1972, 1977\)](#) en littérature et anthropologie, Keynes et Hayek en économie<sup>18</sup> ; pour devenir aujourd'hui un sujet transdisciplinaire extrêmement dynamique, aussi bien en psychologie du développement [[Meltzoff et Prinz \(2002\)](#)], qu'en psychologie sociale [[Bandura et McClelland \(1977\)](#)], en sociologie [[Hedström \(2006\)](#)], en économie [[Durlauf \(2001\)](#); [Weibull \(1995\)](#)], en anthropologie évolutionniste [[Boyd et Richerson \(1988\)](#); [Donald \(1991\)](#)], dans la modélisation multi-agents en sciences sociales [[Conte et Paolucci \(2001\)](#)], en éthologie [[Byrne et Russon \(1998\)](#); [Tomasello \(2003\)](#)], et dans les domaines de la neurologie, la robotique et la philosophie de l'esprit [[Nadel et Decety \(2015\)](#)].

Mais force est de constater qu'en ce qui concerne le domaine de la modélisation de l'évolution culturelle, que ce soit en économie, sociologie ou anthropologie évolution-

18. Voir [Dupuy \(2004a\)](#)

niste, le thème de l'imitation a perdu son unité première. Dans ce regain d'intérêt, chaque approche ne déploie en effet qu'un aspect de l'imitation. Nous prendrons dans ce qui suit l'exemple de deux approches que Jean-Pierre Dupuy a maintes fois confrontées dans ses recherches, à savoir celle de l'économie dite « orthodoxe », et la théorie girardienne du social. La première privilégie la composante rationnelle de l'imitation, dans le prolongement de la notion d'imitation persistante chez Baldwin et de l'imitation logique de Tarde ; alors que la seconde envisage des processus mimétiques opaques aux individus, qui rappelle ce que Baldwin nommait imitation simple et que Tarde rangeait dans les imitations de type « non-logique ».

D'un côté nous trouvons l'idée qu'un individu imite parce qu'il désire, de l'autre celle qu'il désire parce qu'il imite. Ceci se traduit par des choix radicalement différents de modélisation, qui présentent chacun une forme d'incomplétude.

Pour le courant qui s'attache à l'aspect logique et rationnel de l'imitation, celle-ci est un *moyen* au service d'une *fin*. Elle est pensée dans une relation de dépendance par rapport à des principes ou des préférences pré-établis en dehors de tout contexte social : *A* imite *B* s'il estime que ce dernier est particulièrement performant dans l'obtention de quelque chose que lui, *A*, désire déjà en raison de ses propres critères personnels (par exemple, si *A* est cupide, il aura une propension à s'inspirer du comportement des individus qui se seront fait remarquer par leur opulence). Ces principes ou préférences sont une propriété permanente de l'individu *A*, que celui-ci a apporté dans ses valises lorsqu'il est arrivé dans la société.

Il est utile de prendre l'exemple de la théorie économique pour mesurer ce que l'on perd à ne considérer que cet aspect de l'imitation. Les économistes ont longtemps été réticents à l'introduction de la notion d'imitation dans leurs modèles, de peur que par cette brèche ne s'introduise du social susceptible de saboter les fondements individualistes de leurs théories. Lorsque imitation il y a, celle-ci choisit sa direction et ses modèles en fonction des préférences de l'agent, qui sont tenues pour fixes et immuables. Cette fixité des préférences est le socle de l'individualisme méthodologique et a longtemps été considérée comme un axiome, alors qu'elle n'a en fait aucune justification théorique ou empirique. Comme le faisait remarquer Bowles (2001), elle n'est que la marque de l'incomplétude de la théorie économique : « Les économistes, et dans une moindre mesure les autres chercheurs en sciences sociales, ont des réticences à aborder la complexité et le caractère endogène des motivations humaines, non parce que l'on pense que la simplicité comportementale de l'*homo œconomicus* est une représentation adéquate, mais plutôt par manque d'outils conceptuels adéquats et d'information empirique sur le processus de formation des préférences. »

Lorsqu'on sait que tous les modèles micro-économiques s'appuient sur cet axiome pour dériver les processus de décision des agents, c'est tout l'édifice théorique qui se fissure sous cette remarque. L'économiste se retrouve alors dans la position d'un météorologue, qui aurait toute sa vie modélisé la chute des gouttes d'eau, sans s'intéresser à la formation ou au déplacement des nuages.

Devant cette situation, certains économistes ont porté leurs efforts de modélisation sur la représentation des motivations, des identités ou des préférences des agents, de leur hétérogénéité et de leur évolution<sup>19</sup>. La plupart des approches de l'identité s'accordent sur la nécessité d'établir une relation entre formation des traits individuels et formation des groupes sociaux, mais, comme le souligne Davis (2006), chacune présente des

19. Voir par exemple le numéro de la *Revue de Philosophie Économique* consacré à cette question (n° 9, 2004).

défaillances dans le bouclage de l'interdépendance entre la formation des identités des agents, des groupes sociaux, et des réseaux d'interactions qu'ils forment.

Ces recherches ont fait ressortir le besoin de prendre en compte le caractère dynamique et socialement encastré des traits individuels dans la modélisation, ainsi que la difficulté formelle qu'il y a à rendre ceux-ci endogènes, c'est-à-dire résultant d'un processus, et non l'expression de l'*a priori* arbitraire du modélisateur. Cette exigence se trouve formulée de façon concise par le prix Nobel d'économie Vernon Smith (2005) à propos de la théorie de jeux (mes italiques) :

Technically, the issue can be posed as one of asking how most productively to model agent 'types' by *extending game theory so that types are an integral part of its predictive content*, rather than merely imported as an ex post technical explanation of experimental results. For example, moves can signal types, and effect decision, which explains why game-form matters, and why opportunity cost—payoffs available, but foregone—can effect outcomes. These considerations must be part of the internal structure of the theory such that outcomes become predictions conditional on the elementary characteristics of *players who read each other's intentions*. If such a program were successful, many of the basic results in game theory would become irrelevant or special cases of the extended theory.

Le premier passage en italiques souligne la nécessité de rendre endogènes les types d'agents ; le second donne une piste pour opérer cette endogénéisation : considérer des agents capables de raisonner sur le niveau procédural et non pas uniquement sur le niveau comportemental.

Comme le soulignait déjà Vernon Smith dans son discours de réception du prix Nobel [Smith (2003)], "*the point that needs emphasis is that it is easy to go from "types" (traditionally utility or beliefs about states) to game theoretic choice ; the hard part is to relate "types" to characteristics of the individual's memory-sensory system.*" C'est cette "*hard part*" que je me suis attelée à poursuivre dans mes recherches.

La question de la modélisation de l'imitation dans son aspect logique et rationnel rencontre des limites qui touchent aux fondements mêmes de la science économique, dans sa conception de ce que sont les agents et leurs préférences. Nous pourrions espérer voir ces limites dépassées en nous tournant vers des penseurs qui ont mené leur travaux en dehors de cette discipline, et ont fait de l'imitation la clé de voûte de leur théorie. C'est le cas, par exemple, de René Girard, pour qui la question n'est pas d'identifier les préférences ou les désirs qui orientent l'imitation, mais de partir de l'hypothèse qu'un individu désire *parce qu'il imite*. La conséquence de cette simple hypothèse est l'introduction dans les dynamiques sociales d'un phénomène de rétroaction positive, qui transforme de manière contingente de l'indifférencié en différencié.

Cette rétroaction positive, dont le pouvoir morphogénétique a été maintes fois souligné par des auteurs comme Jean-Pierre Dupuy ou André Orléans [Orléans (2011)], a pour origine ce que Girard a appelé la *médiation double*. La valeur des choses se lit dans les désirs des autres. Parmi tous les objets désirables, c'est parce *A* croit que *B* désire *O* qu'il va également le désirer. Mais symétriquement, le désir de *B* sera aiguisé par le désir naissant de *A*, confirmant ou renforçant la croyance initiale de ce dernier. Qu'importe alors que cette croyance fut vraie ou fausse, le jeu de miroirs des imitations va sélectionner *O*, de manière potentiellement contingente, comme point focal des désirs de *A* et *B*, et justifier *a posteriori* les croyances de chacun.

Comme l'a souligné à plusieurs reprises Jean-Pierre Dupuy, ce principe permet de rendre compte, de manière stylisée, de l'émergence d'une polarisation des désirs humains, sans aucune structure préalable et par un processus intrinsèquement social. L'individu, dans ses désirs et ses aspirations, n'existe pas en dehors de l'interaction avec autrui. Ce sont les effets de l'imitation, démultipliés par le kaléidoscope des interactions, qui font émerger les différents « types » d'agents. Cet exemple est une illustration de la démarche propre à l'*individualisme méthodologique complexe*.

La théorie girardienne et les formalisations qu'ont pu en faire Jean-Pierre Dupuy ou André Orléans comportent des subtilités qui mériteraient de plus amples développements<sup>20</sup>. On peut cependant émettre des réserves quant à la portée de cette formalisation. La théorie girardienne oublie en effet que, hormis les situations très particulières où il y a indifférenciation totale des individus (par exemple les scènes primitives évoquées par Girard), le fait qu'un individu ait telle ou telle structure sur ses désirs ou ses préférences, contribue à déterminer la manière dont il sera sujet à influence mimétique. Selon qu'on soit fan de boxe, de danse ou de tricot, on ne sera pas influencé par les mêmes personnes. Or chez Girard, la forme de l'imitation, c'est-à-dire la manière dont un individu choisit ses modèles, est une propriété universelle qui éclipse les caractéristiques propres des individus. Les différents degrés de la notion girardienne de *médiateur*<sup>21</sup> (interne et externe) n'y remédient pas. Traduite dans le langage formel de la modélisation, l'imitation girardienne reste un conformisme « généralisé », dont l'originalité est de se situer sur le plan des désirs (ou des préférences) des agents. Et c'est ainsi que tous les modèles qui s'appuient sur cette théorie la représentent<sup>22</sup>.

#### 4.2.4 Repenser l'imitation

Les deux approches de l'imitation que nous venons d'examiner rompent le lien entre individu et collectif en donnant trop, ou pas assez, de poids aux caractéristiques individuelles. Mais il suffirait de les mettre en vis-à-vis pour s'apercevoir que la solution est peut-être sous nos yeux. *J'imite parce que je désire, je désire parce que j'imite* : n'est-ce pas là un de ces processus circulaires tels qu'on peut les rencontrer dans les systèmes autopoïétiques, systèmes qui engendrent leur propres espaces de significations ? Car il suffit de reconnaître que l'individu n'est pas le siège passif de processus mimétiques, qu'il a une certaine capacité de rétention sémiotique, pour proposer un compromis entre l'approche girardienne et l'approche rationnelle de l'imitation.

Si nous prolongeons les réflexions de Baldwin et de Tarde dans une perspective de modélisation, il nous faut envisager une complexification des formes d'imitation à mesure que s'élabore la personnalité d'un individu, et reconnaître dans la multiplicité des formes que prend l'imitation, l'effet d'une différenciation qui opère, en premier lieu, au niveau des désirs et des préférences d'une population.

Du point de vue dynamique, il s'agit donc de modéliser un processus au cours duquel, alors qu'un individu se forge une personnalité, notamment par influences mimétiques, la forme de l'imitation évolue, modifiant d'emblée l'ensemble des modèles ou « médiateurs » potentiels. L'ensemble des règles culturelles, des normes ou des croyances qu'un individu acquiert au cours de sa vie, sont autant d'éléments qui vont contraindre les formes

20. Voir par exemple Dupuy (1990)

21. « Médiateur » est le terme employé par Girard pour désigner la personne qui sert de modèle aux désirs d'un individu.

22. Pour un des premiers modèles d'inspiration girardienne en économie, voir Orléans (1985).

de ses imitations futures. Il y a alors une relation de dépendance réciproque entre la personnalité d'un individu et les processus mimétiques susceptibles de faire évoluer cette personnalité. Traduit dans le langage de la théorie des jeux, cela revient à exprimer une co-détermination entre le « type » d'un agent et les règles d'imitation qu'il met en œuvre.

Posé de cette manière, le problème n'est pas tant d'arriver à concevoir l'émergence d'une organisation sociale ou économique à partir d'interactions entre agents pleinement constitués antérieurement et extérieurement à cette organisation, comme le fait la théorie économique, mais d'arriver à penser ensemble l'individu et le social « comme se créant mutuellement, se définissant et se contenant l'un l'autre » [Dupuy (1990), op. cit., p. 229] ou, pour reprendre une formulation plus technique de Dupuy (2004b), remplacer une figure de point fixe exogène par une figure de point fixe endogène.

*L'une de mes plus importantes contributions à la modélisation en sciences humaines et sociales est la formalisation de cette relation de double dépendance, entre formation des types d'agents et comportements mimétiques, dans un cadre théorique que j'ai appelé les jeux metamimétiques.*

Les jeux metamimétiques reposent sur une formalisation d'un principe d'imitation qui intègre la dimension réflexive et procédurale de la cognition humaine, dimension que, soit dit en passant, l'espèce humaine est quasiment la seule à avoir développée<sup>23</sup>.

Pour reprendre le vocabulaire de la théorie de jeux, ce formalisme permet d'exprimer, dans un même système d'équations d'évolution, des utilités ou préférences incommensurables. Les distributions possibles des préférences ou des types d'agents font intrinsèquement partie des solutions de ce système d'équations et deviennent par là même endogènes (contrairement aux modèles standards en économie et autres modèles formels des sciences humaines et sociales, où les préférences, traits ou désirs des agents sont des paramètres exogènes des modèles). D'un point de vue conceptuel, le recours aux jeux metamimétiques abouti donc à ce que Vernon Smith appelait de ses vœux, et dont la conséquence est de proposer une modélisation des dynamiques sociales suivant un mode endogène.

*Comme le préconisait Smith, j'ai montré que le formalisme des jeux metamimétiques est une généralisation de la théorie des jeux évolutionnaires, auquel correspond une nouvelle notion d'équilibre et d'attracteur, les équilibres (resp. attracteurs) metamimétiques, ou équilibres (resp. attracteurs) stochastiquement contrefactuellement stables [Chavalarias (2004, 2006b)].*

Dans un langage plus technique, mais en restant générique, soit  $P$  une population d'agents et  $\Gamma_A$  l'ensemble des informations sur les autres agents de  $P$  auxquelles  $A$  a accès (son « point de vue sur  $P$  »). Soit  $\nu_A(B, \Gamma_A) \in \mathfrak{R}$  l'utilité attribuée par l'agent  $A$  lorsqu'il s'imagine être contrefactuellement à la place de  $B$ , étant donnée sa connaissance de  $\Gamma_A$ . La comparaison interpersonnelle effectuée par  $A$  s'exprime en comparant  $\nu_A(A, \Gamma_A)$  et  $\nu_A(B, \Gamma_A)$ .

On définit alors une fonction de satisfaction relative :

$$S(A, \Gamma_A) = \begin{cases} 1, & \text{si } \nu_A(B_{max}, \Gamma_A) - \nu_A(B_{min}, \Gamma_A) < \epsilon \\ \frac{\nu_A(A, \Gamma_A) - \nu_A(B_{min}, \Gamma_A)}{\nu_A(B_{max}, \Gamma_A) - \nu_A(B_{min}, \Gamma_A)}, & \text{Sinon} \end{cases} \quad (1)$$

23. Quelques éléments de raisonnement réflexif ont été mis en évidence chez les chimpanzés et les dauphins (par exemple des expériences révèlent que les chimpanzés se reconnaissent dans le miroir ou peuvent anticiper qu'ils ne sauront pas réaliser une tâche), mais, jusqu'à preuve du contraire, ils sont relativement peu développés.

où  $B_{max} = \max_{B \in \Gamma_A} \nu_A(B, \Gamma_A)$ ,  $B_{min} = \min_{B \in \Gamma_A} \nu_A(B, \Gamma_A)$  et où  $\epsilon$  est un seuil en dessous duquel  $A$  ne peut pas identifier de différence d'utilité<sup>24</sup>.

$S(A, \Gamma_A) < 1$  signifie que  $A$  n'est pas satisfait et va subir de manière accrue l'influence sociale d'agents dont il estime que leur situation est plus en adéquation avec ses propres préférences (les  $B$  tels que  $\nu_A(B, \Gamma_A) > \nu_A(A, \Gamma_A)$ ). Cette influence sociale va conduire  $A$  à réviser ses préférences autant que ses comportements (méta-imitation) (voir Chavalarias (2006b) pour les détails).

La dynamique metamimétique d'une population d'agents est définie par la succession des comparaisons inter-personnelles, qui donnent lieu à une évolution des comportements et des préférences.

On définit la satisfaction globale de la population par :

$$\Psi = \sum_{A \in P} S(A, \Gamma_A)$$

Les équilibres de cette dynamique (*équilibres métamimétiques*) sont les états tels que  $\Psi = 1$ . La dynamique est hors équilibre dès lors que  $\Psi < 1$ .

Il est important de remarquer que la fonction  $\Psi$  agrège des utilités qui ne sont pas nécessairement commensurables, aucune hypothèse particulière n'est fait en ce sens sur les fonctions  $\nu$ . Celles-ci sont des fonctions  $\nu : \{A, \Gamma_A\} \rightarrow \{v_1, \dots, v_{|\Gamma_A|}\}$  de  $P \otimes \Gamma_* \rightarrow \mathbb{R}^{|\Gamma_*|}$  où  $P$  est l'ensemble des agents,  $\Gamma_*$  l'ensemble des voisinages sociaux possibles (ensemble d'agents avec leurs caractéristiques observables) et  $\mathbb{R}^{|\Gamma_*|}$  l'ensemble des n-tuples réels de taille inférieure ou égale à la taille de la population<sup>25</sup>.

Les travaux de modélisation et de simulation informatique de ces jeux métamimétiques [Chavalarias (2008c)] montrent que l'introduction de cette modalité particulière de comportement mimétique engendre des dynamiques sur l'espace des types d'agents qui aboutissent à une structuration hétérogène de la population autour d'un ensemble d'états bien déterminés (attracteurs). Ces dynamiques sont générées par des collectifs d'individus qui, considérés de manière isolée, en dehors de tout historique d'interactions sociales, sont définis par des ensembles de « *virtualités élémentaires* »<sup>26</sup>, large spectre de types possibles mais indéterminés, dont aucun n'est privilégié a priori. Ce n'est que lorsque l'agent est plongé dans un contexte d'interactions mimétiques qu'il actualise une instantiation particulière de l'un des « types » possibles d'agents.

J'ai illustré ce phénomène dans le cadre de modèles multi-agents de dilemmes sociaux, qui donnent à voir des dynamiques telles que la composition des types d'agents, sous des contraintes définies localement par les types eux-mêmes, forme au niveau collectif un « tissu social artificiel », qui exhibe des sous-structures cohérente et interprétables en terme de capacités cognitives des agents (cf. fig. 28) : larges groupes d'agents conformistes, anti-conformistes isolés, agents matérialistes plus mobiles que la moyenne et moins satisfaits que la moyenne, etc.

Dans une telle approche, non seulement les structures collectives sont des propriétés émergentes de la dynamique sociale, mais la structure cognitive interne de chaque agent l'est également. Individus et collectif sont co-construits et peuvent être vus comme des

24. Cette fonction de satisfaction est une première approximation, différentes personnes pouvant avoir différentes manières de normaliser cette comparaison sociale en fonction de leur caractère. Ce type de données relève des travaux de psychologie sociale.

25. Chaque agent  $A$  établit ses comparaisons sociales sur la base des agents qu'il peut observer dans son voisinage social, leur nombre, noté  $|\Gamma_A|$ , vérifiant nécessairement  $|\Gamma_A| \leq |P|$ .

26. Expression employée par Tarde (1898) dans *Les Lois Sociales - Esquisse d'une sociologie*



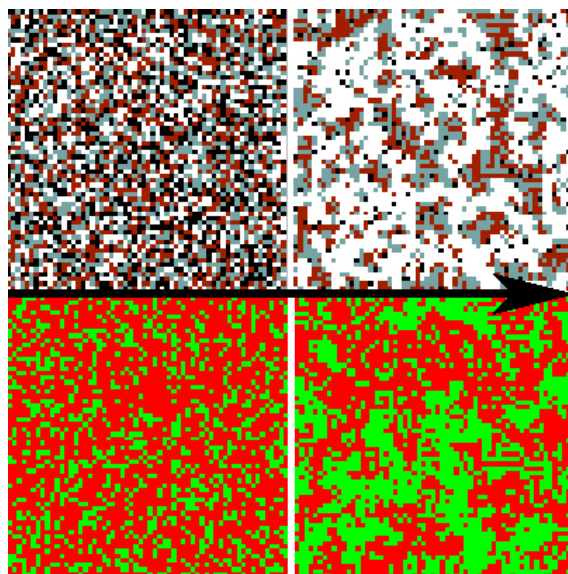


FIGURE 28 – Exemple d'évolution d'une population metamimetique jouant à un dilemme du prisonnier spatial et répété à partir d'un état aléatoire. (le temps coule de gauche à droite). Les agents peuvent se réaliser dans quatre types distincts : conformistes, anti-conformistes, et agents aux préférences matérialistes (mini et maxi). Devinez qui est qui. En haut, la distribution des types d'agents, en bas, la distribution des comportements ( $C$  et  $D$ ). L'état de droite est un attracteur metamimetique obtenu après une vingtaine de cycles d'imitation.

présentations parallèles d'un même phénomène, le système social artificiel, qui n'existe pas en dehors de son actualisation. Les individus ne sont plus constitués extérieurement au social mais élaborent leurs préférences au cours du chemin qu'ils suivent ; le social ne précède pas la mise en relation des individus mais est contenu en chaque individu par l'ensemble des virtualités élémentaires qu'il possède<sup>27</sup>. Par ailleurs, les jeux metamimétiques donnent à observer différentes échelles de temps (temps individuel, le temps des groupes et le temps du collectifs) et permettent de questionner leur articulation.

Ce type de systèmes présente des configurations particulières, équilibres ou attracteurs, remarquables par leur stabilité relative. Il est intéressant de noter que contrairement à la définition standard d'états stables en économie, qui porte sur les comportements des agents étant donnée leurs préférences, lorsque l'on rend endogène la *formation des préférences*, la distribution des préférences est elle-même l'un des résultats de la dynamique qu'elle définit. Dans le cas de modèles mimétiques, cette dynamique peut être qualifiée de metamimétique. La condition de stabilité d'un tel système est alors qu'il existe des états de la population dont la distribution des préférences soit telle, qu'elle soit à la fois un point fixe de la dynamique mimétique qu'elle engendre, et définisse une dynamique sur les comportements pour laquelle il existe également un point fixe, ce que nous avons appelé *états contrefactuellement stables*<sup>28</sup>.

Plus généralement, ce formalisme permet d'endogénéiser un grand nombre de paramètres des modèles de dynamiques sociales, avec pour objectif que toute variable qui ne peut être rendue endogène via ce formalisme (comme par exemple un taux d'erreur dans l'inférence d'un trait ou l'exécution d'un comportement) doit pouvoir être estimée expéri-

27. Ce qui n'est pas incompatible avec le fait que les individus puissent se constituer au sein d'une organisation sociale pré-établie.

28. Cf. Chavalarias 2006.

mentalement ou empiriquement (par des expériences en sciences cognitives ou des études sociologiques ou économiques).

Ce formalisme s'est montré très fécond pour conceptualiser des processus de morphogenèse dans les systèmes sociaux. Je n'ai cessé d'y travailler depuis mon doctorat, en montrant en quoi il proposait une méthode générique pour endogénéiser les différents traits caractérisants les agents : préférences, échelles de temps, mobilités, etc. [Chavalarias (2006b, 2008c, 2013); Chavalarias et Batta (2013)]. J'ai également montré que les jeux metamimétiques peuvent être vus comme la traduction dans un langage de modélisation de certains concepts clés de la sociologie de Gabriel Tarde [Chavalarias (2007b)].

Comme Smith l'appelait de ses vœux, l'approche que j'ai proposée permet de poser des questions qui ne peuvent être formulées dans le cadre classique de la théorie des jeux : comment la distribution des préférences des agents est influencée par des paramètres tels que : la topologie des interactions, les institutions auxquelles ils sont confrontés, les changements exogènes (ex : la force du dilemme social auxquels ils sont soumis - on peut penser par exemple aux effets négatifs du réchauffement climatique sur certaines activités humaines). Les jeux metamimétiques permettent de proposer une modélisation idéal-typique de la société sans avoir recours à l'hypothèse de l'objectivité de la valeur (individualisme méthodologique) ou des déterminismes sociaux (holisme).

Le cadre des jeux metamimétiques m'a permis de formuler un certain nombre de prédictions et conjectures qui sont empiriquement vérifiables :

1. Effets de facteurs environnementaux sur la distributions des types d'agents (par ex. force du dilemme social),
2. Multiplicité des attracteurs des dynamiques sociales et caractère hétérogène de ces attracteurs (au sens de la mixité des types),
3. Hiérarchisation des échelles de temps en fonction de la proximité à la praxis des traits dont elles dépendent,
4. Effet de la topologie du réseau d'interaction et de la fiabilité des informations échangées sur la distribution des types d'agents,
5. Degré moyen de satisfaction des agents en fonction du type de préférences (matérialistes ou sociales).

Ce formalisme donne une nouvelle perspective à la question de l'évaluation du bien-être d'une population, abordée par exemple par l'économie du bien-être (*welfare economics*). Que veut dire en effet "maximiser" ou "optimiser" le bien-être d'une population alors que la distribution des préférences peut changer sous l'effet des mesures prises ? N'est-il pas justement dans l'essence des mesures politiques que d'anticiper les transformations des préférences qu'elles induisent ?

Le point 4 pose par ailleurs la question du couplage entre nos infrastructures technologiques et l'évolution culturelle, couplage dont j'ai montré qu'il était plus que jamais d'actualité dans Chavalarias (2016).

Toutes ces questions sont actuellement au cœur de mes recherches et leurs réponses requièrent d'adopter une approche fondamentalement interdisciplinaire. Elles bénéficient par ailleurs d'un nouvel élan apporté par la révolution des méga-données, qui ouvre des perspectives de corroboration/réfutation empirique de ce type de travaux. Alors que jusqu'à très récemment, il était difficilement envisageable de disposer d'une observation empirique de l'évolution des préférences à l'échelle d'une population ou avec une granularité temporelle suffisante au niveau individuel, ceci devient concevable avec l'exploitation des

traces numériques laissées par toutes nos activités en ligne (cf. section 3), qui permettent d'inférer, à qui sait capter ces traces et les analyser, les goûts, les opinions politiques, les catégories socio-professionnelles, les pratiques sportives, les orientations sexuelles, etc. des individus.

Il suffit par exemple d'aller visiter un site comme MySpace pour se rendre compte de la mosaïque humaine que ces nouvelles technologies donnent à voir (cf. figure 29).

Les grands acteurs du web, qui disposent de jeux de données colossaux, font la démonstration quotidienne que cette perspective est en train de devenir une réalité (voir aussi Chavalarias (2016)). Un groupe tel que Twitter, nous donne un exemple de ce qu'il est possible de faire (cf. Fig. 30 et 31). En recoupant les informations données par ses partenaires commerciaux, Twitter établit une vision très détaillée des personnalités de ses utilisateurs. Cette analyse d'audience est probablement complétée par tout ce qu'il est possible de faire avec de la fouille de donnée textuelle et de l'analyse des graphes : détermination de l'humeur, des réseaux familiaux, professionnels et amicaux, des opinions politiques, des centres d'intérêt, etc. Twitter ayant ces données en temps réel et sur de longues périodes, cette entreprise est en mesure d'analyser les évolutions de comportements et de personnalité de ses usagers.

La recherche académique ne peut espérer avoir accès à des données aussi exhaustives et confidentielles pour étudier l'évolution culturelle. Elle n'est cependant pas en reste. Il y aura à l'avenir suffisamment de données pour pouvoir faire des études longitudinales - sociologiques, économiques, historiques, etc. - sans précédent, qui permettront d'étudier en détails l'évolution culturelle et la structuration sociale de nos sociétés.

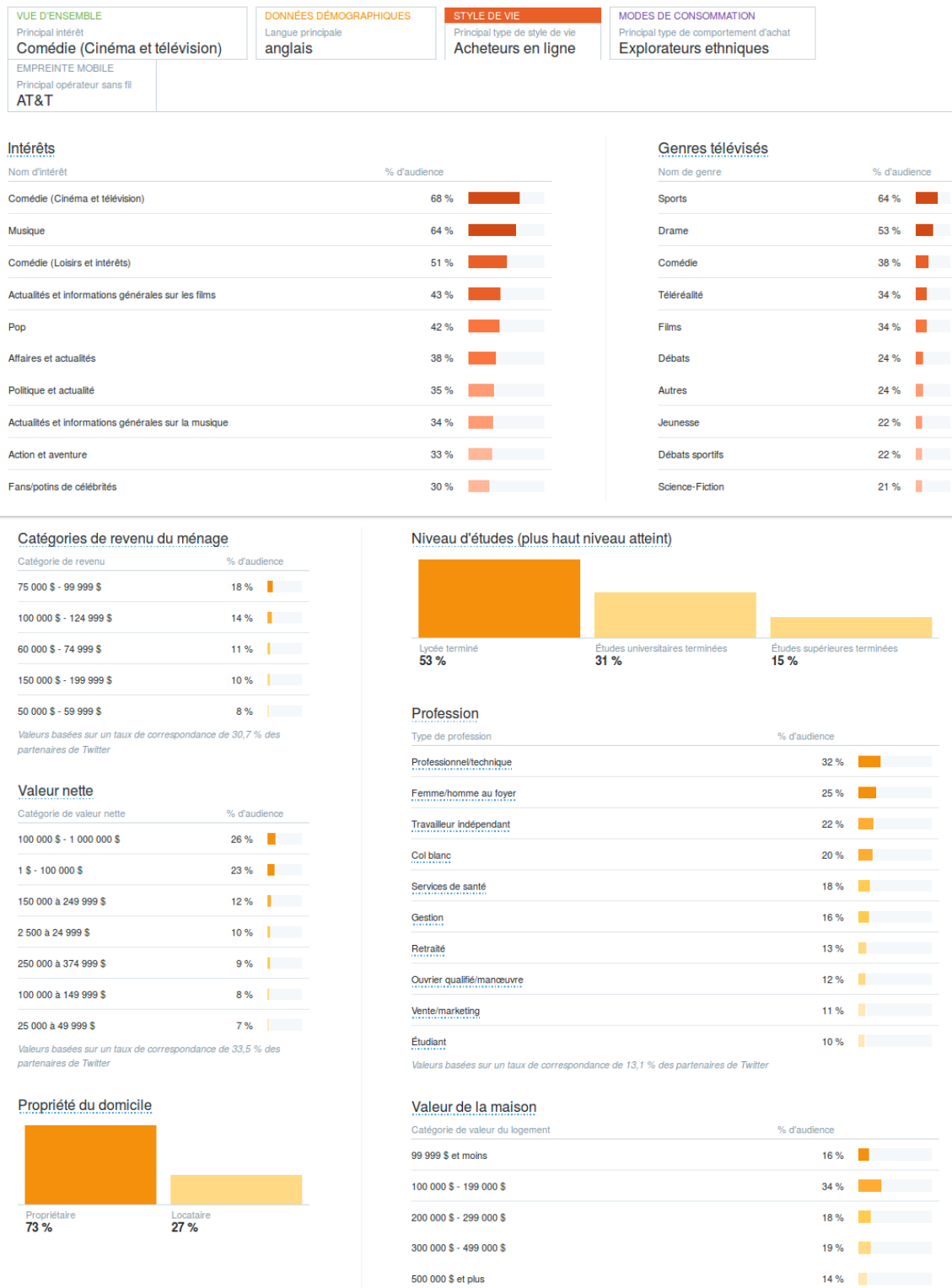
Ces jeux de données de terrains seront complétés par des jeux de données provenant de nouveaux types d'expérience à grande échelle. La recherche académique est en effet en train de saisir les opportunités offertes par les technologies mobiles, et notamment la démocratisation des smartphones qui les rendent ubiquitaires, pour déployer à moindre coût des expériences en ligne sur de très large cohortes, dans des domaines aussi divers que la psychologie, l'économie ou les marchés culturels. Un certain nombre de travaux se sont aventurés sur ce terrain, laissant entrevoir la possibilité de mesurer et d'analyser, avec des approches interdisciplinaires, la formation des préférences et l'effet des interactions sociales sur celles-ci. Les biens culturels ont été l'un des terrains d'étude avec des méthodes expérimentales [Ormerod et Glass (2010); Salganik et collab. (2006)] ou d'enquête [Dolfsma (2004)], comme je l'ai analysé dans [Chavalarias (2016)].

### 4.3 Éthique des TIC

Mes recherches peuvent aboutir à des résultats et des outils qui, utilisés dans un cadre autre qu'académique (sur des bases de données différentes avec des objectifs différents) peuvent soulever un certain nombre de questions éthiques. Alors que les guru du high-tech nous promettent monts et merveilles avec le développement des technologies de l'information et de la communication (TIC), des mega-données et du calcul haute performance, plusieurs affaires récentes, dont les révélations Snowden, ont mis en lumière les risques et dérives possibles de l'utilisation de ces nouvelles technologies. Pour cette raison, il me semble important que les chercheurs impliqués dans ce domaine mènent également une réflexion critique sur l'usage des TIC. Après un premier article dans un ouvrage collectif [Chavalarias (2012)], les révélations Snowden m'ont incité à poursuivre cette réflexion dans un second article paru dans la revue *Big Data and Society* [Chavalarias (2016)].



FIGURE 29 – Les quelques 50M d'utilisateurs de MySpace peuvent donner libre cours à leur imagination dans le choix de leurs avatars et identités virtuelles, choix qui se font souvent en groupe, comme on peut l'observer sur ces captures d'écran (2007) qui montrent les meilleurs "amis" de certains utilisateurs. L'un d'eux a été artificiellement substitué. La facilité à l'identifier est un bon indice de la forte structuration de ces groupes autour de centres d'intérêts bien identifiés.



**FIGURE 30 – Exemple d’analyse d’audience de Twitter.** En recoupant avec les informations données par ses partenaires commerciaux, Twitter établi une vision très détaillée des personnalités de ses clients. Cette analyse d’audience complète tout ce qu’il est possible de faire avec de la fouille de donnée textuelle et de l’analyse des graphe et qui permet de déterminer les réseaux familiaux, professionnels et amicaux, les opinions politiques, etc. De plus, elle prend en compte les dynamiques temporelles, ce qui permet à Twitter d’analyser les évolutions de comportements et de personnalité.

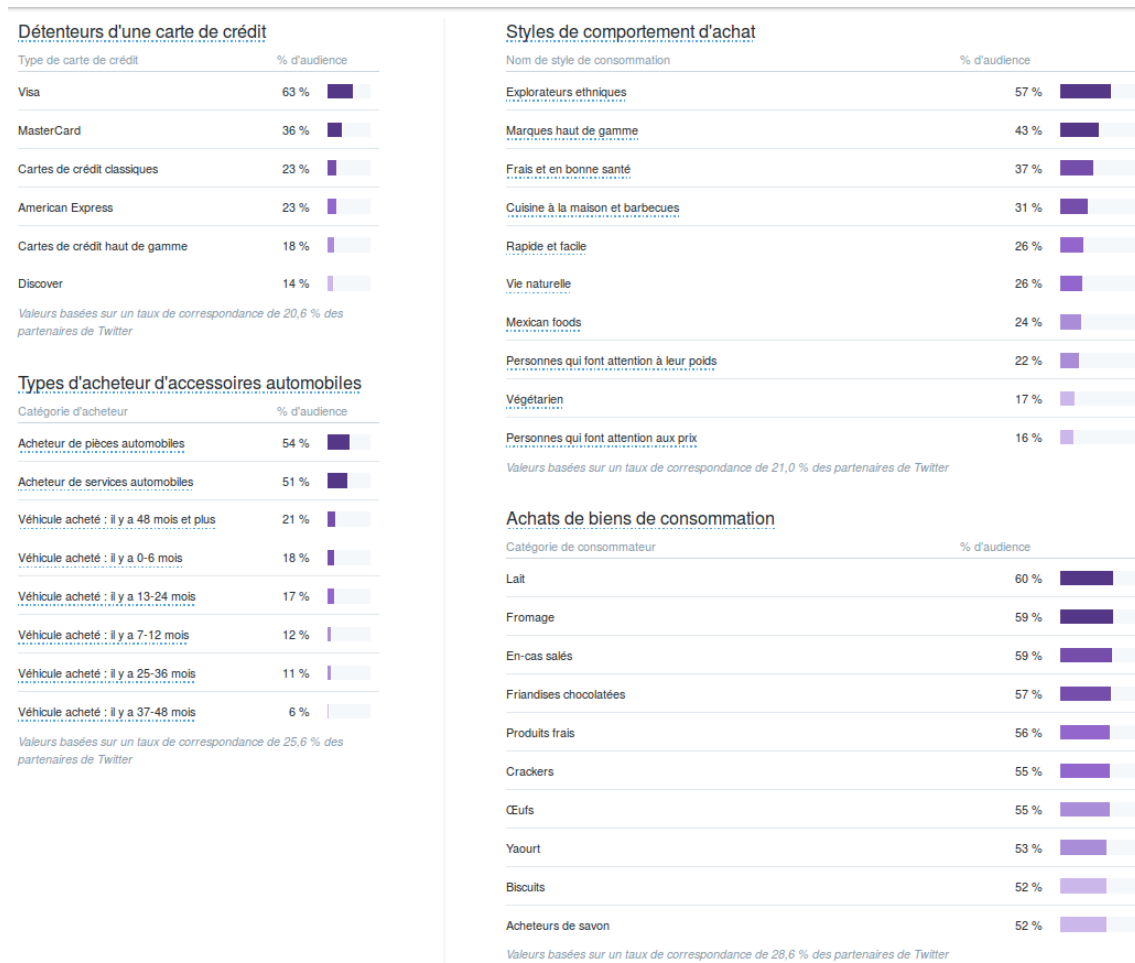


FIGURE 31 – Exemple d'analyse d'audience de Twitter. Idem qu'en fig.30.

## PUBLICATIONS ET TRAVAUX ASSOCIÉS

- Chavalarias, D., 2004. *Metadynamiques en cognition sociale - Quelle définition de meilleure est la meilleure ?* Ecole Polytechnique, Paris.
- Chavalarias, D., 2006. *Metamimetic Games : Modeling Metadynamics in Social Cognition*. *Journal of Artificial Societies and Social Simulation* 9, 5.
- Chavalarias, David, 2007. *L'articulation individu/collectif dans les sciences des systèmes complexes : quels apports pour la sociologie ?* Sociétés.
- Chavalarias, D., 2008. *Cooperation as an outcome of a social differentiation process in a metamimetic game*, in : Edmonds, B., Hernández-Iglesias, C., Troitzsch, K.G. (Eds.), *Social Simulation : Technologies, Advances and New Discoveries*. Information Science Reference, Hershey, PA.
- Chavalarias, David, 2009. *Désir mimétique et imitation rationnelle. Vers un individualisme méthodologique complexe*, in : Anspach, M.R., Centre culturel international de Cerisy-la-Salle (Eds.), Jean-Pierre Dupuy, dans *l'œil du cyclone : colloque de Cerisy*. Carnets nord, Paris.
- Chavalarias, David, 2012. *When preferences heterogeneity and agents types become endogenous : the example of metamimetic games*. 17th Annual Workshop on Economic Heterogeneous Interacting Agents (WEHIA), Paris.
- Chavalarias, David, Batta, Erasmo, 2013. *Modeling endogenous preferences for heterogeneous agents adressing "the hard part,"* in : European Conference on Complex Systems (ECCS'13). European Conference on Complex Systems, Barcelona.
- Chavalarias, D., 2012. *La société (re)commandée : de la conjecture de von Foerster aux sciences sociales prédictives*, in : Chardel, A., Gossart, C., Reber, B. (Eds.), *Conflits Des Interprétations Dans La Société de L'information*.
- D. Chavalarias, « *The unlikely encounter between von Foerster and Snowden - When second-order cybernetics sheds light on societal impacts of Big Data* », *Big Data and Society* (à paraître, accepté à *Big Data and Society*, Sage Journals).

---

## 5 Conclusions

---

En ce début de XXI<sup>ème</sup> siècle, les sciences sociales subissent une transformation probablement comparable à la révolution qu'a connu la biologie dans la deuxième moitié du XX<sup>ème</sup> siècle.

L'avènement de la biologie moléculaire a ouvert la voie à une description fine des processus maintenant l'intégrité de la cellule et assurant son articulation avec l'organisme. Dans le même temps, de nouveaux équipements (micro-array, microscopie confocale, IRM, etc.) et le développement des capacités de calcul ont permis d'observer, notamment *in vivo*, les organismes de manière multi-échelle en espace (échelle moléculaire, cellulaire, organes) et en temps (de l'expression des gènes à l'ontogenèse). Il en a résulté des découvertes majeures, une accélération du rythme de publication (multiplié par trois en vingt ans), et surtout une approche beaucoup plus intégrée du vivant, les théories proposant de prendre en compte différentes échelles d'espace et de temps ayant acquis de solides bases empiriques.

De nombreux signes indiquent que les sciences sociales connaissent une évolution analogue. Depuis les années 1990, les développements des sciences cognitives et de la psychologie sociale ont apporté des percées très importantes dans la compréhension de l'homme et de ses rapports à son environnement social.

Parallèlement, la migration de pans entiers d'activités sociales vers des supports numériques (communications, commerce, réseaux sociaux, productions culturelles, etc.) contribue au phénomène des méga-données, et nous donne à observer et analyser *in vivo* les systèmes sociaux d'une façon inédite.

Enfin, la démocratisation des TIC et leur portabilité a favorisé le développement de méthodes expérimentales en économie, sociologie ou psychologie (notamment grâce à des laboratoires virtuels) qui permettent de déployer à moindre coût, des dispositifs expérimentaux pouvant traiter de très larges cohortes, ce dans des environnements très variés (lieu de travail, écoles, hôpitaux, pays en voie de développement, etc.).

Du fait des possibilités étendues d'observation et d'expérimentation en espace et en temps, cette évolution apporte un nouveau regard sur des questions fondamentales des sciences sociales - par exemple l'articulation individu - collectif, comportements - institutions. Elle devrait également permettre de stimuler un grand nombre de domaines de recherche en rendant les prédictions de certaines théories empiriquement vérifiables via l'acquisition de nouveaux types de données au niveau individuel et collectif.

Ceci étant, nous pouvons attendre de cette évolution des sciences sociales une nouvelle compréhension des phénomènes sociaux collectifs et des dynamiques sociales, qu'ils concernent les comportements, les sentiments, les représentations, la mémoire ou la formation des valeurs socio-culturelles.

Une partie de mes travaux de ces deux dernières années [Chavalarias (2016)] ont consisté à démontrer la convergence et la consistance de ces nouvelles approches expérimentales et empiriques avec les approches modélisatrices.

La recherche française ne pourrait prendre ce tournant des sciences sociales qu'avec le soutien de centres de recherche interdisciplinaires capables d'apporter aux chercheurs les moyens et les compétences couvrant l'ensemble de ces besoins. C'est la raison pour laquelle j'ai investi une partie importante de mon activité de chercheur dans le développement et la pérennisation d'un tel centre, l'Institut des Systèmes Complexes de Paris



Île-de-France, dont j'ai été directeur adjoint de 2010 à 2012, et dont j'ai pris la direction en 2013 (cf. section 7). Cet institut a catalysé de nombreux projets dans ce domaine (ANR, ERC, Equipex, etc.) et j'ai réussi depuis ma prise de fonctions à y constituer une équipe d'ingénieurs à même d'accompagner les chercheurs dans le développement de ces nouvelles méthodologies pour les sciences sociales.

Lauréat du programme *Emergence(s)* (Mairie de Paris) en 2013, j'ai également lancé à l'ISC-PIF un axe de recherche ayant pour objectif de développer des méthodologies et des ressources autour du troisième pilier de ce tournant des sciences sociales : l'expérimentation et la collecte distribuée de données.

Le fil directeur de mes recherches a été de faire converger ces trois approches (reconstruction phénoménologique, reconstruction théorique et dispositifs expérimentaux pour la collecte de données) vers une compréhension multi-échelle des dynamiques sociales et des processus de morphogenèse - différenciation sociale, formation des réseaux sociaux, émergence de normes, formation des valeurs socio-culturelles. Le champ est immense et les sciences humaines et sociales ne sont qu'au début de leur mue.

Sans attendre, reprenons à notre compte le mot d'Herbert Simon : il y a d'un côté les sciences "dures", et de l'autre les sciences "très dures", les sciences humaines et sociales ; et cette difficulté est extrêmement stimulante.

# Bibliographie

- Ahn, T., D. Chavalarias et A. Lazar. 2002, «Endogenous Network Formation and Evolution of Preferences», dans *Proceedings of the Workshop on Agent Simulation : Applications, Models, and Tools*, Argonne, IL : Argonne National Laboratory.
- Armatte, M. 2005, «La notion de modèle dans les sciences sociales : anciennes et nouvelles significations», *Mathématiques et sciences humaines. Mathematics and social sciences*, n° 172. 00031.
- Atlan, H. 1999, *Les étincelles de hasard*, La librairie du XXe siècle, Seuil, Paris, ISBN 978-2-02-025248-5. 00068.
- Atlan, H. 2014, *La science est-elle inhumaine ? : essai sur la libre nécessité*, Bayard, Paris, ISBN 978-2-227-48766-6. 00053.
- Baldwin, J. M. -. 1897, *Le développement mental chez l'enfant et dans la race / par James Mark Baldwin,.... ; traduit de l'anglais par M. Nourry ; et précédé d'une préface de M. Léon Marillier,...*
- Baldwin, J. M. 1894, «Imitation : a Chapter in the Natural History of Consciousness», *Mind*, vol. III, n° 9, doi :10.1093/mind/III.9.26, p. 26–55, ISSN 0026-4423, 1460-2113.
- Bandura, A. et D. C. McClelland. 1977, «Social learning theory», .
- Barreteau, O., M. Antona, P. D'Aquino, S. Aubert, S. Boissau, F. Bousquet, W. Daré, M. Etienne, C. Le Page, R. Mathevet, G. Trébuil et J. Weber. 2003, «Our Companion Modelling Approach», *Journal of Artificial Societies and Social Simulation*.
- Benecke, A. 2006, «Chromatin code, local non-equilibrium dynamics, and the emergence of transcription regulatory programs», *The European Physical Journal E*, vol. 19, n° 3, doi : 10.1140/epje/i2005-10068-8, p. 353–366, ISSN 1292-8941, 1292-895X. 00038.
- Bengio, Y., A. Courville et P. Vincent. 2013, «Representation Learning : A Review and New Perspectives», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, n° 8, doi :10.1109/TPAMI.2013.50, p. 1798–1828, ISSN 0162-8828, 2160-9292.
- Berestycki, H., J. P. Nadal et N. Rodríguez. 2015, «A model of riots dynamics : Shocks, diffusion and thresholds», *Networks and Heterogeneous Media*, vol. 10, n° 3, doi :10.3934/nhm.2015.10.443, p. 443–475, ISSN 1556-1801. 00001.
- Bonabeau, E., M. Dorigo et G. Theraulaz. 1999, *Swarm intelligence : from natural to artificial isystems*, Oxford University Press, New York, ISBN 978-0-19-513158-1 978-0-19-513159-8.
- Bonabeau, E. et G. Theraulaz. 1994, *Intelligence collective*, Hermes Paris, France.
- Boorman, S. A. et H. C. White. 1976, «Social Structure from Multiple Networks. II. Role Structures», *American Journal of Sociology*, vol. 81, n° 6, p. 1384–1446, ISSN 0002-9602.
- Borghesi, C., L. Hernández, R. Louf et F. Caparros. 2013, «Universal size effects for populations in group-outcome decision-making problems», *Physical Review E*, vol. 88, n° 6, doi : 10.1103/PhysRevE.88.062813, ISSN 1539-3755, 1550-2376.
- Börner, K. 2011, «Plug-and-play macroscopes», *Communications of the ACM*, vol. 54, n° 3, doi :10.1145/1897852.1897871, p. 60, ISSN 00010782. 00059.
- Börner, K. 2015, *Atlas of knowledge : anyone can map*, The MIT Press, Cambridge, Massachusetts, ISBN 978-0-262-02881-3.
- Borner, K., C. M. Chen et K. W. Boyack. 2003, «Visualizing knowledge domains», *Annual Review of Information Science and Technology*, vol. 37, doi :10.1002/aris.1440370106, p. 179–255, ISSN 0066-4200. 00868 WOS :000179918000006.
- Bourgine, P., D. Chavalarias et C. Cohen-Boulakia, éd.. 2008, *Déterminismes et complexités : du physique à l'éthique : autour d'Henri Atlan*, Recherches, La Découverte, Paris, ISBN 978-2-7071-5090-5.
- Bourgine, P. et J. Stewart. 2004, «Autopoiesis and Cognition», *Artificial Life*, vol. 10, n° 3, doi :10.1162/1064546041255557, p. 327–345, ISSN 1064-5462, 1530-9185.
- Bowles, S. 2001, «Individual Interactions, Group Conflicts, and the Evolution of Preferences», dans *Social dynamics*, édité par S. N. Durlauf, n° 4 dans Economic learning and social evolution, Brookings Inst. Press [u.a.], Washington, DC, ISBN 978-0-262-54176-3 978-0-262-04186-7.
- Bowles, S. et H. Gintis. 2011, *A cooperative species : human reciprocity and its evolution*, Princeton University Press, Princeton, ISBN 978-0-691-15125-0.
- Boyd, R. et P. J. Richerson. 1988, *Culture and the evolutionary process*, paperback ed éd., University of Chicago Press, Chicago (u.a.), ISBN 978-0-226-06933-3.
- Boyer, R. 2007, «Peut-on mettre en évidence des lois en économie ? : Le point de vue d'un macroéconomiste régulationniste», *Economies et sociétés*, p. 1799–1812.
- Braam, R. R., H. F. Moed et A. F. J. van Raan. 1991, «Mapping of science by combined co-citation and word analysis. I. Structural aspects», *Journal of the American Society for Information Science*, vol. 42, n° 4, doi :10.1002/(SICI)1097-4571(199105)42:4<233::

- AID-AS11>3.0.CO;2-I, p. 233–251, ISSN 0002-8231, 1097-4571. 00310.
- Breiger, R. L. 1974, «The Duality of Persons and Groups», *Social Forces*, vol. 53, n° 2, doi :10.1093/sf/53.2.181, p. 181–190, ISSN 0037-7732, 1534-7605.
- Bressloff, P. C., J. D. Cowan, M. Golubitsky, P. J. Thomas et M. C. Wiener. 2002, «What Geometric Visual Hallucinations Tell Us about the Visual Cortex», *Neural Computation*, vol. 14, n° 3, doi :10.1162/089976602317250861, p. 473–491, ISSN 0899-7667, 1530-888X. 00160.
- Byrne, R. W. et A. E. Russon. 1998, «Learning by imitation : a hierarchical approach», *Behavioral and brain sciences*, vol. 21, n° 05, p. 667–684.
- Callon, M., J. Courtial, W. Turner et S. Bauin. 1983, «From Translations to Problematic Networks - an Introduction to Co-Word Analysis», *Social Science Information Sur Les Sciences Sociales*, vol. 22, n° 2, doi :10.1177/053901883022002003, p. 191–235, ISSN 0539-0184. 00548 WOS :A1983QW30800003.
- Callon, M., J. P. Courtial et F. Laville. 1991a, «Co-word analysis as a tool for describing the network of interactions between basic and technological research : The case of polymer chemistry», *Scientometrics*, vol. 22, n° 1, doi :10.1007/BF02019280, p. 155–205, ISSN 0138-9130, 1588-2861. 00490.
- Callon, M., J.-P. Courtial et W. Turner. 1991b, «La méthode Leximappe : un outil pour l'analyse stratégique du développement scientifique et technique», *Gestion de la recherche. Nouveaux problèmes, nouveaux outils, Bruxelles : De Boeck*, p. 207–277. 00022.
- Cavalli-Sforza, L. L. et M. W. Feldman. 1981, *Cultural transmission and evolution : a quantitative approach*, n° 16 dans *Monographs in population biology*, Princeton University Press, Princeton, N.J., ISBN 978-0-691-08280-6 978-0-691-08283-7. 03382.
- Chateauraynaud, F. 2003, *Prospéro : une technologie littéraire pour les sciences humaines*, CNRS.
- Chavalarias, D. 2002, «Emergence of cooperation and selection of interactions», dans *International Symposium On Complexity*, Shanghai. [Http ://arxiv.org/abs/nlin/0205067](http://arxiv.org/abs/nlin/0205067).
- Chavalarias, D. 2004, *Metadynamiques en cognition sociale - Quelle définition de meilleure est la meilleure ?*, thèse de doctorat, Ecole Polytechnique, Paris.
- Chavalarias, D. 2006a, «Cooperation : popularity or cliquishness?», dans *Proceedings of the 6th European Conference on Complex Systems*, édité par J. Jost, F. Reed-Tsochas et P. Schuster, Oxford. 00000.
- Chavalarias, D. 2006b, «Metamimetic Games : Modeling Metadynamics in Social Cognition», *Journal of Artificial Societies and Social Simulation*, vol. 9, n° 2, p. 5, ISSN 1460-7425.
- Chavalarias, D. 2007a, «La part mimétique des dynamiques de cognition sociale : clé pour penser l'auto-transformation du social», *Nouvelles Perspectives en Sciences Sociales*.
- Chavalarias, D. 2007b, «L'articulation individu/collectif dans les sciences des systèmes complexes : quels apports pour la sociologie?», *Sociétés*.
- Chavalarias, D. 2008a, «Biologie», dans *Déterminismes et complexités : du physique à l'éthique : autour d'Henri Atlan*, édité par P. Bourguine, D. Chavalarias et C. Cohen-Boulakia, Recherches, La Découverte, Paris, ISBN 978-2-7071-5090-5, p. 85–89.
- Chavalarias, D. 2008b, «Cognition, Conscience et Liberté», dans *Déterminismes et complexités : du physique à l'éthique : autour d'Henri Atlan*, édité par P. Bourguine, D. Chavalarias et C. Cohen-Boulakia, Recherches, La Découverte, Paris, ISBN 978-2-7071-5090-5, p. 209–2014. 00000.
- Chavalarias, D. 2008c, «Cooperation as an outcome of a social differentiation process in a metamimetic game», dans *Social simulation : technologies, advances and new discoveries*, édité par B. Edmonds, C. Hernández-Iglesias et K. G. Troitzsch, Information Science Reference, Hershey, PA, ISBN 978-1-59904-522-1 978-1-59904-524-5.
- Chavalarias, D. 2008d, «Physiques et formalismes», dans *Déterminismes et complexités : du physique à l'éthique : autour d'Henri Atlan*, édité par P. Bourguine, D. Chavalarias et C. Cohen-Boulakia, Recherches, La Découverte, Paris, ISBN 978-2-7071-5090-5, p. 17–23.
- Chavalarias, D. 2008e, «Politique et Ethique», dans *Déterminismes et complexités : du physique à l'éthique : autour d'Henri Atlan*, édité par P. Bourguine, D. Chavalarias et C. Cohen-Boulakia, Recherches, La Découverte, Paris, ISBN 978-2-7071-5090-5, p. 313–319.
- Chavalarias, D. 2009, «Désir mimétique et imitation rationnelle. Vers un individualisme méthodologique complexe», dans *Jean-Pierre Dupuy, dans l'œil du cyclone : colloque de Cerisy*, édité par M. R. Anspach et Centre culturel international de Cerisy-la-Salle, Carnets nord, Paris, ISBN 978-2-35536-020-6.
- Chavalarias, D. 2012, «La société (re)commandée : de la conjecture de von Foerster aux sciences sociales prédictives», dans *Conflits des interprétations dans la société de l'information*, édité par A. Chardel, C. Gossart et B. Reber, librairie lavoisier éd..
- Chavalarias, D. 2013, «Modeling endogenous heterogeneity in social systems», .

- Chavalarias, D. 2016, «The unlikely encounter between von Foerster and Snowden : When second-order cybernetics sheds light on societal impacts of Big Data», *Big Data & Society*, vol. 3, n° 1, doi :10.1177/2053951715621086, ISSN 2053-9517.
- Chavalarias, D. forthcoming, «What's wrong with Science ? Modeling the collective discovery processes with the Nobel Game», *Scientometrics*.
- Chavalarias, D. et E. Batta. 2013, «Modeling endogenous preferences for heterogeneous agents addressing "the hard part"», dans *European Conference on Complex Systems (ECCS'13)*, Barcelona.
- Chavalarias, D., P. Bourguine, E. Perrier, F. Amblard, F. Arlabosse, P. Auger, J.-B. Baillon, O. Barreteau, P. Baudot, E. Bouchaud, S. Ben Amor, H. Berry, C. Bertelle, M. Berthod, G. Beslon, G. Biroli, D. Bonamy, D. Bourcier, N. Brodu, M. Bui, Y. Burnod, B. Chapron, C. Christophe, B. Clément, J.-L. Coatrieux, J.-P. Cointet, V. Dagrain, K. Dauchot, O. Dauchot, F. Daviaud, S. De Monte, G. Deffuant, P. Degond, J.-P. Delahaye, R. Doursat, F. D'Ovidio, M. A. Dubois, B. Dubruelle, M. Dutreix, R. Faivre, E. Farge, P. Flandrin, S. Franceschelli, C. Gaucherel, J.-P. Gaudin, M. Ghil, J.-L. Giavitto, F. Ginelli, V. Ginot, F. Houllier, B. Hubert, P. Jensen, L. Jullien, Z. Kapoula, D. Krob, F. Ladiou, G. Lang, C. Lavelle, A. Le Bivic, J.-P. Leca, C. Lecerc, P. Legrain, D. L'Hôte, M. Loireau, J.-F. Mangin, O. Monga, M. Morvan, J.-P. Muller, I. Negrutiu, N. Peyreiras, D. Pumain, O. Radulescu, J. Sallantin, E. Sanchez, D. Schertzer, M. Schoenauer, M. Sebag, E. Simonet, A. Six, F. Tarissan et P. Vincent. 2009, «French Roadmap for complex Systems 2008-2009», cahier de recherche, Cargèse.
- Chavalarias, D., S. Castillo et M. Panahi. 2015, «Le Tweeposcope Climatique. Une représentation collective des enjeux autour du climat», *Lettre de l'INSHS*, vol. 38, p. 39-42. [Http ://iscpif.fr/wp-content/uploads/2015/11/Tweeposcope\\_INSHS\\_lettre\\_infoinshs-38.pdf](http://iscpif.fr/wp-content/uploads/2015/11/Tweeposcope_INSHS_lettre_infoinshs-38.pdf)
- Chavalarias, D., S. Charron, V. Gardelle et P. Bourguine. 2006, «NOBEL, Le jeu de la découverte scientifique, Une approche analytique, expérimentale et computationnelle», dans *conférence Modélisation, Optimisation et Simulation des Systèmes, Défis et Opportunités (MOSIM'06)*, Rabbat.
- Chavalarias, D. et J.-P. Cointet. 2008, «Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study», *Scientometrics*, vol. 75, n° 1, doi :10.1007/s11192-007-1825-6, p. 37-50, ISSN 0138-9130, 1588-2861.
- Chavalarias, D. et J.-P. Cointet. 2009, «The Reconstruction of Science Phylogeny», *arXiv :0904.3154 [physics]*.
- Chavalarias, D. et J.-P. Cointet. 2013a, «Phylomemetic patterns in science evolution—the rise and fall of scientific fields», *PLoS one*, vol. 8, n° 2, p. e54847.
- Chavalarias, D. et J.-P. Cointet. 2013b, «Science Phylomemy», [Http ://scimaps.org/maps/map/science\\_phylomemy\\_159/](http://scimaps.org/maps/map/science_phylomemy_159/).
- Chavalarias, D., J.-P. Cointet et C. Roth. 2010, «Étude portant sur un outil de veille dynamique sur l'attente sociétale en matière de recherche et d'innovation», Recherche et développement, Ministère de l'Enseignement Supérieur et de la Recherche, Paris. [Http ://tina.iscpif.fr/veille/mesr](http://tina.iscpif.fr/veille/mesr).
- Chavalarias, D. et V. Gardelle. 2008, «Social simulation of collective a discovery process : the Nobel Game», dans *5th conference of the European Social Simulation Association*, Brescia, Italy.
- Chavalarias, D. et J. P. Ioannidis. 2010, «Science mapping analysis characterizes 235 biases in biomedical research», *Journal of Clinical Epidemiology*, vol. 63, n° 11, doi :10.1016/j.jclinepi.2009.12.011, p. 1205-1215, ISSN 08954356.
- Chavalarias, D., C. Jean-Philippe, T. K. Duong, L. Cornilleau, L. Villard, C. Roth et T. Savy. 2011a, «Streams of Media Issues - Monitoring World Food Security», dans *UN Global Pulse Symposium*, United Nations, New York. [Http ://pulseweb.cortext.net](http://pulseweb.cortext.net).
- Chavalarias, D., J.-B. Leenart et L. Hernández. 2014, «Municipale 2014, ne vous absteniez pas de savoir», <http://municipales.iscpif.fr/>. [Http ://municipales.iscpif.fr/](http://municipales.iscpif.fr/).
- Chavalarias, D., E. Showk et J. Bilcke. 2011b, «Tinasoft», [Http ://tinasoft.eu](http://tinasoft.eu).
- Chavalarias, D., J. D. Wallach, A. H. T. Li et J. P. A. Ioannidis. 2016, «Evolution of Reporting P Values in the Biomedical Literature, 1990-2015», *JAMA*, vol. 315, n° 11, doi :10.1001/jama.2016.1952, p. 1141, ISSN 0098-7484.
- Chavalarias, D. 1998, «La thèse de Popper est-elle réfutable», Master thesis, Ecole Polytechnique, Paris.
- Church, A. 1936a, «A note on the Entscheidungsproblem», *The journal of symbolic logic*, vol. 1, n° 01, p. 40-41. 00779.
- Church, A. 1936b, «An unsolvable problem of elementary number theory», *American journal of mathematics*, p. 345-363. 01695.
- Cointet, J.-P. et D. Chavalarias. 2008, «Multi-level Science mapping with asymmetric co-occurrence analysis : Methodology and case study», *Networks and Heterogeneous Media*, p. 267-276.
- Conte, R. et M. Paolucci. 2001, «Intelligent Social Learning», <http://jasss.soc.surrey.ac.uk/4/1/3.html>.
- d'Alembert, J. I. R. 1751, *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers*, vol. Tome 1. 00952 = "dictionary".
- Danermark, B. 2003, «Review of Re-Thinking Science : Knowledge and the Public in an Age of Uncertainty», *Acta Sociologica*, vol. 46, n° 2, p. 166-169, ISSN 0001-6993. 00000.

- Davis, J. B. 2006, «Social identity strategies in recent economics», *Journal of Economic Methodology*, vol. 13, n° 3, doi : 10.1080/13501780600908168, p. 371–390, ISSN 1350-178X, 1469-9427.
- de Rosnay, J. 1975, *Le macroscope : vers une vision globale*, Editions du Seuil, Paris, ISBN 978-2-02-002818-9.
- Deffuant, G., A. Banos, D. Chavalarias, C. Bertelle, N. Brodu, P. Jensen, A. Lesne, J.-P. Müller, É. Perrier et F. Varenne. 2015, «Visions de la complexité. Le démon de Laplace dans tous ses états», *Natures Sciences Sociétés*, vol. 23, n° 1, doi :10.1051/nss/2015007, p. 42–53, ISSN 1240-1307, 1765-2979.
- Delanoë, A., D. Chavalarias, S. Castillo, R. Loth et M. Rodic. 2015, «Gargantext», [Http://gargantext.org](http://gargantext.org).
- Dieckmann, U., R. Law, J. A. J. Metz et International Institute for Applied Systems Analysis, éd.. 2000, *The geometry of ecological interactions : simplifying spatial complexity*, Cambridge studies in adaptive dynamics, Cambridge University Press, Cambridge, U.K. ; New York, NY, USA, ISBN 978-0-521-64294-1. 00481.
- Dolfsma, W. 2004, *Institutional Economics and the Formation of Preferences : The Advent of Pop Music*, Edward Elgar Publishing, ISBN 978-1-84542-072-7.
- Donald, M. 1991, *Origins of the modern mind : three stages in the evolution of culture and cognition*, Harvard University Press, Cambridge, Mass, ISBN 978-0-674-64483-0.
- Dupuy, J.-P. 1990, *Ordres et désordres enquête sur un nouveau paradigme*, Ed. du Seuil, Paris, ISBN 978-2-02-010923-9.
- Dupuy, J. P. 1992, *Introduction aux sciences sociales : logique des phénomènes collectifs*, Ellipses, Edition Marketing, Paris, ISBN 978-2-7298-9226-5.
- Dupuy, J.-P. 1995, «Sciences Sociales et Sciences Cognitives», .
- Dupuy, J.-P. 2004a, «Economics as Symptom», dans *Transforming economics : perspectives on the critical realist project*, édité par P. Lewis, Economics as social theory, Routledge, London ; New York, ISBN 978-0-415-36966-4 978-0-415-36967-1.
- Dupuy, J.-P. 2004b, «Vers l'unité des sciences sociales autour de l'individualisme méthodologique complexe», *Revue du MAUSS*, vol. 24, n° 2, doi :10.3917/rdm.024.0310, p. 310, ISSN 1247-4819, 1776-3053.
- Durlauf, S. N., éd.. 2001, *Social dynamics*, n° 4 dans Economic learning and social evolution, Brookings Inst. Press [u.a.], Washington, DC, ISBN 978-0-262-54176-3 978-0-262-04186-7.
- Edmond, B., Gilbert, N., Ahrweiler, P. et Scharnhorst, A. 2011, «Simulating the Social Processes of Science», *JASSS*.
- Epstein, R. et R. E. Robertson. 2015, «The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections», *Proceedings of the National Academy of Sciences*, vol. 112, n° 33, doi :10.1073/pnas.1419828112, p. E4512–E4521, ISSN 0027-8424, 1091-6490.
- Farge, E. 2003, «Mechanical Induction of Twist in the Drosophila Foregut/Stomodaeal Primordium», *Current Biology*, vol. 13, n° 16, doi :10.1016/S0960-9822(03)00576-1, p. 1365–1377, ISSN 09609822. 00313.
- Fehr, E. et U. Fischbacher. 2003, «The nature of human altruism», *Nature*, vol. 425, n° 6960, doi :10.1038/nature02043, p. 785–791, ISSN 0028-0836, 1476-4679.
- Freeman, L. C. 1977, «A Set of Measures of Centrality Based on Betweenness», *Sociometry*, vol. 40, n° 1, p. 35–41.
- Garfield, E. 1955, «Citation Indexes for Science : A New Dimension in Documentation through Association of Ideas», *Science*, vol. 122, n° 3159, doi :10.1126/science.122.3159.108, p. 108–111, ISSN 0036-8075, 1095-9203. 01870.
- Garfield, E. 1974, «Isi Is Studying Structure of Science Through Co-Citation Analysis», *Current Contents*, , n° 7, p. 5–6. 00012 WOS :A1974ZZ06900001.
- Gilbert, G. N. et S. Woolgar. 1974, «The Quantitative Study of Science : An Examination of the Literature», *Science Studies*, vol. 4, n° 3, p. 279–294, ISSN 0036-8539.
- Gilbert, N. 1997, «Gilbert : A Simulation of the Structure of Academic Science», *Sociological Research Online*.
- Girard, R. 1972, *La violence et le sacré*, Pluriel, Paris, ISBN 978-2-8185-0206-8 978-2-01-278897-8.
- Girard, R. 1977, *Mensonge romantique et vérité romanesque*, grasset éd..
- Granovetter, M. 1978, «Threshold models of collective behavior», *American journal of sociology*, p. 1420–1443. 03545.
- Grim, P. 1994, «Undecidability in the Spatialized Prisoner's Dilemma : Some Philosophical Implications», dans *IEEE International Conference on Computational Intelligence*. 00005.
- Guerrien, B. 2004, «Y a-t-il une science économique ?», *L'Economie Politique*, p. 97–109.
- He, X. et J. Zhang. 2009, «On the Growth of Scientific Knowledge : Yeast Biology as a Case Study», *PLoS Computational Biology*, vol. 5, n° 3, doi :10.1371/journal.pcbi.1000320, p. e1000 320, ISSN 1553-7358.
- Hedström, P. 2006, «Explaining Social Change : an Analytical Approach», *Papers : revista de sociologia*, , n° 80, p. 73–95, ISSN 2013-9004.

- Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, N. S. Henrich, K. Hill, F. Gil-White, M. Gurven, F. W. Marlowe, J. Q. Patton et D. Tracer. 2005, «“Economic man” in cross-cultural perspective : Behavioral experiments in 15 small-scale societies», *Behavioral and Brain Sciences*, vol. 28, n° 06, doi :10.1017/S0140525X05000142, ISSN 0140-525X, 1469-1825.
- Horst, U., A. Kirman et M. Teschl. 2006, «Changing identity : the emergence of social groups», *GREQAM Document de Travail*, vol. 51.
- Hull, D. L. 1988, *Science as a process : an evolutionary account of the social and conceptual development of science*, Science and its conceptual foundations, University of Chicago Press, Chicago, ISBN 978-0-226-36050-8.
- Hutson, V. et G. Vickers. 2000, «Methods for Reaction-Diffusion Models», dans *The geometry of ecological interactions : simplifying spatial complexity*, édité par U. Dieckmann, R. Law, J. A. J. Metz et International Institute for Applied Systems Analysis, Cambridge studies in adaptive dynamics, Cambridge University Press, Cambridge, U.K. ; New York, NY, USA, ISBN 978-0-521-64294-1, p. 456-481. 00000.
- Ioannidis, J. P. A. 2005, «Why Most Published Research Findings Are False», *PLoS Medicine*, vol. 2, n° 8, doi :10.1371/journal.pmed.0020124, p. e124, ISSN 1549-1676.
- Kanizsa, G. 1979, *Organization in vision : Essays on Gestalt perception*, Praeger Publishers. 01644.
- Kessler, M. 1963, «Bibliographic Coupling Between Scientific Papers», *American Documentation*, vol. 14, n° 1, doi : 10.1002/asi.5090140103, p. 10-&, ISSN 0096-946X. 01169 WOS :A19632554A00006.
- Kitchin, R. 2014, «Big Data, new epistemologies and paradigm shifts», *Big Data & Society*, vol. 1, n° 1, doi :10.1177/2053951714528481, ISSN 2053-9517. 00067.
- Kuhn, T. S. 1962, *The structure of scientific revolutions*, 3<sup>e</sup> éd., University of Chicago Press, Chicago, IL, ISBN 978-0-226-45807-6 978-0-226-45808-3.
- Lahire, B. 2005, «Sociologie, psychologie et sociologie psychologique», dans *Psychologie sociale et communication*, édité par B. Orfali, n° 41 dans Hermès, CNRS Éd, Paris, ISBN 978-2-271-06313-7, p. 151-157. 00000.
- Latour, B., P. Jensen, T. Venturini, S. Grauwin et D. Boullier. 2012, «‘The whole is always smaller than its parts’ - a digital test of Gabriel Tarde’s monads : ‘The whole is always smaller than its parts’», *The British Journal of Sociology*, vol. 63, n° 4, doi :10.1111/j.1468-4446.2012.01428.x, p. 590-615, ISSN 00071315. 00165.
- Latour, B. et S. Woolgar. 1986, *Laboratory life : the construction of scientific facts*, Princeton University Press, Princeton, N.J, ISBN 978-0-691-02832-3 978-0-691-09418-2. 00067.
- Lazega, E. 2016, «Synchronization Costs in the Organizational Society : Intermediary Relational Infrastructures in the Dynamics of Multilevel Networks», dans *Multilevel Network Analysis for the Social Sciences*, édité par E. Lazega et T. A. Snijders, Springer International Publishing, Cham, ISBN 978-3-319-24518-8 978-3-319-24520-1, p. 47-77.
- Lazega, E. et T. A. Snijders, éd.. 2016, *Multilevel Network Analysis for the Social Sciences : Theory, Methods and Applications*, 1<sup>re</sup> éd., Methodos Series 12, Springer International Publishing, ISBN 978-3-319-24518-8 978-3-319-24520-1.
- Lesne, A. 2008, «Déterminisme et aléatoire dans les systèmes complexes : un faux débat ?», dans *Déterminismes et complexités : du physique à l'éthique : autour d'Henri Atlan*, édité par P. Bourguin, D. Chavalarias et C. Cohen-Boulakia, Recherches, La Découverte, Paris, ISBN 978-2-7071-5090-5, p. 85-89. 00000.
- Lesne, A. et J. M. Victor. 2006, «Chromatin fiber functional organization : Some plausible models», *The European Physical Journal E*, vol. 19, n° 3, doi :10.1140/epje/i2005-10050-6, p. 279-290, ISSN 1292-8941, 1292-895X. 00048.
- Lestel, D. 2001, *Les origines animales de la culture*, Flammarion, Paris, ISBN 978-2-08-212527-7.
- Lorenz, E. 1972, «The Butterfly effect», 00000.
- McMullin, B. et F. Varela. 1997, «Rediscovering Computational Autoipoiesis», 00000.
- Meeker, M. 2015, «Key Internet Trends», cahier de recherche, KPCB. [Http ://www.kpcb.com/internet-trends](http://www.kpcb.com/internet-trends).
- Meltzoff, A. N. et W. Prinz, éd.. 2002, *The imitative mind : development, evolution, and brain bases*, Cambridge studies in cognitive perceptual development, Cambridge University Press, Cambridge, U.K. ; New York, ISBN 978-0-521-80685-5.
- Mill, J. S. 1843, *Système de logique déductive et inductive. : exposé des principes de la preuve et des méthodes de recherche scientifique*, *Classiques des sciences sociales.*, vol. VI, J.-M. Tremblay, Chicoutimi, ISBN 978-1-55442-219-7. 00019.
- Morgan, M. S. et M. Morrison, éd.. 1999, *Models as mediators : perspectives on natural and social sciences*, Ideas in context, Cambridge University Press, Cambridge ; New York, ISBN 978-0-521-65097-7 978-0-521-65571-2.
- Morin, E. et E. Morin. 1977, *La nature de la nature*, n° t. 1 dans *His La Méthode*, Seuil, Paris, ISBN 978-2-02-004634-3. 02341.
- Nadel, J. et J. Decety. 2015, «Imiter pour découvrir l'humain», .
- Nowak, M. A., S. Bonhoeffer et R. M. May. 1994, «Spatial games and the maintenance of cooperation», *Proceedings of the National Academy of Sciences*, vol. 91, n° 11, p. 4877-4881, ISSN 0027-8424, 1091-6490.

- Nowak, M. A. et R. M. May. 1993, «The Spatial Dilemmas of Evolution», *International Journal of Bifurcation and Chaos*, vol. 03, n° 01, doi :10.1142/S0218127493000040, p. 35–78, ISSN 0218-1274, 1793-6551. 00607.
- Nowak, M. A. et K. Sigmund. 2000, «Games on Grids», dans *The geometry of ecological interactions : simplifying spatial complexity*, édité par U. Dieckmann, R. Law, J. A. J. Metz et International Institute for Applied Systems Analysis, Cambridge studies in adaptive dynamics, Cambridge University Press, Cambridge, U.K. ; New York, NY, USA, ISBN 978-0-521-64294-1, p. 135–153. 00481.
- Olivier, N., M. A. Luengo-Oroz, L. Duloquin, E. Faure, T. Savy, I. Veilleux, X. Solinas, D. Debarre, P. Bourguine, A. Santos, N. Peyrieras et E. Beaurepaire. 2010, «Cell Lineage Reconstruction of Early Zebrafish Embryos Using Label-Free Nonlinear Microscopy», *Science*, vol. 329, n° 5994, doi :10.1126/science.1189428, p. 967–971, ISSN 0036-8075, 1095-9203.
- Orléan, A. 2011, *L'empire de la valeur : refonder l'économie*, Couleur des idées, Éditions du Seuil, Paris, ISBN 978-2-02-105437-8.
- Orléans, A. 1985, «Monnaie et spéculation mimétique», dans *Violence et vérité : autour de René Girard*, édité par P. Dumouchel et Centre culturel international de Cerisy-la-Salle, B. Grasset, Paris, ISBN 978-2-246-33721-8, p. 147–158.
- Ormerod, P. et K. Glass. 2010, «Predictability in an 'Unpredictable' Artificial Cultural Market», dans *Advances in Social Computing*, vol. 6007, édité par D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, S.-K. Chai, J. J. Salerno et P. L. Mabry, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-642-12078-7 978-3-642-12079-4, p. 354–359.
- Palla, G., A.-L. Barabási et T. Vicsek. 2007, «Quantifying social group evolution», *Nature*, vol. 446, n° 7136, doi :10.1038/nature05670, p. 664–667, ISSN 0028-0836. 01073.
- Palla, G., I. Derényi, I. Farkas et T. Vicsek. 2005, «Uncovering the overlapping community structure of complex networks in nature and society», *Nature*, vol. 435, n° 7043, doi : 10.1038/nature03607, p. 814–818, ISSN 0028-0836, 1476-4679. 03286.
- Payette, N. 2012, «Agent-Based Models of Science», dans *Models of Science Dynamics*, édité par A. Scharnhorst, K. Börner et P. van den Besselaar, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-642-23067-7 978-3-642-23068-4, p. 127–157.
- Peters, H. P. F. et A. F. J. van Raan. 1993, «Co-word-based science maps of chemical engineering. Part II : Representations by combined clustering and multidimensional scaling», *Research Policy*, vol. 22, n° 1, doi :10.1016/0048-7333(93)90032-D, p. 47–71, ISSN 0048-7333. 00061.
- Petitot, J. 2003, «The neurogeometry of pinwheels as a sub-Riemannian contact structure», *Journal of Physiology-Paris*, vol. 97, n° 2-3, doi :10.1016/j.jphysparis.2003.10.010, p. 265–309, ISSN 09284257.
- Petitot, J. et Y. Tondut. 1999, «Vers une neurogéométrie. Fibrations corticales, structures de contact et contours subjectifs modaux», *Mathématiques informatique et sciences humaines*, n° 145, p. 5–102. 00087.
- Piaget, J., J. Piaget, J. Piaget, S. Psychologist, J. Piaget et S. Psychologue. 1945, *La formation du symbole chez l'enfant : imitation, jeu et rêve, image et représentation*, Delachaux et Niestlé Paris.
- Popper, K. R. 2002, *The logic of scientific discovery*, Routledge, London ; New York, ISBN 978-0-203-99462-7 978-0-415-27843-0 978-0-415-27844-7.
- Popper, K. R. 2006, *Conjectures et réfutations*, Payot, Paris, ISBN 978-2-228-90058-4.
- Roth, C. et J.-P. Cointet. 2010, «Social and semantic coevolution in knowledge networks», *Social Networks*, vol. 32, n° 1, doi : 10.1016/j.socnet.2009.04.005, p. 16–29, ISSN 0378-8733.
- Rudder, C. 2014, *Dataclism : who we are when we think no one's looking*, first edition éd., Crown Publishers, New York, ISBN 978-0-385-34737-2.
- Salganik, M. J., P. Sheridan Dodds et D. Watts. 2006, «Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market», *Science*, vol. 311, n° 5762, doi :10.1126/science.1121066, p. 854–856, ISSN 0036-8075, 1095-9203.
- Schelling, T. C. 1978, *Micromotives and macrobehavior*, 1<sup>re</sup> éd., Fels lectures on public policy analysis, Norton, New York, ISBN 978-0-393-05701-0 978-0-393-09009-3.
- Sengpiel, F., P. Stawinski et T. Bonhoeffer. 1999, «Influence of experience on orientation maps in cat visual cortex», *Nature neuroscience*, vol. 2, n° 8, p. 727–732. 00182.
- Shahaf, D., J. Yang, C. Suen, J. Jacobs, H. Wang et J. Leskovec. 2013, «Information cartography : creating zoomable, large-scale maps of information», dans *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, p. 1097–1105.
- Smith, V. L. 2003, «Constructivist and Ecological Rationality in Economics», dans *Les Prix Nobel. The Nobel Prizes 2002*, p. 502–561.
- Smith, V. L. 2005, «Behavioral economics research and the foundations of economics», *The Journal of Socio-Economics*, vol. 34, n° 2, doi :10.1016/j.soec.2004.09.003, p. 135–150, ISSN 1053-5357.

- Stanley, E. A., D. Ashlock et M. D. Smucker. 1995, «Iterated prisoner's dilemma with choice and refusal of partners : Evolutionary results», dans *Advances in Artificial Life*, vol. 929, édité par G. Goos, J. Hartmanis, J. Leeuwen, J. G. Carbonell, J. Siekmann, F. Morán, A. Moreno, J. J. Merelo et P. Chacón, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-59496-3 978-3-540-49286-3, p. 490–502.
- Stewart, J. 2008, «La vie existe-t-elle ?», dans *Déterminismes et complexités : du physique à l'éthique : autour d'Henri Atlan*, édité par P. Bourguin, D. Chavalarias et C. Cohen-Boulakia, Recherches, La Découverte, Paris, ISBN 978-2-7071-5090-5, p. 85–89. 00000.
- Stryker, S. et P. J. Burke. 2000, «The Past, Present, and Future of an Identity Theory», *Social Psychology Quarterly*, vol. 63, n° 4, doi :10.2307/2695840, p. 284–297, ISSN 0190-2725.
- Tarde, G. -. 1898, *Les lois sociales : esquisse d'une sociologie / par G. Tarde*.
- Tarde, G. 1890, *Les lois de l'imitation : étude sociologique / par G. Tarde*, F. Alcan (Paris).
- Tomasello, M. 2003, *The cultural origins of human cognition*, 4<sup>e</sup> éd., Harvard Univ. Press, Cambridge, Mass., ISBN 978-0-674-00582-2.
- Trabal, P. 2004, «Francis CHATEAURAYNAUD, Prospero. Une technologie littéraire pour les sciences humaines», *Questions de communication*, vol. n° 5, n° 1, ISSN 1633-5961.
- Turing, A. M. 1936, «On computable numbers, with an application to the Entscheidungsproblem», *J. of Math*, vol. 58, n° 345-363, p. 5. 08058.
- Ullmo, J. 1969, *Le profit*, Dunod, Paris.
- von Foerster, H. 1981, *Observing Systems*, intersystems publications éd..
- von Hayek, F. A., W. W. Bartley, S. Kresge et F. A. Hayek. 1978, *The fatal conceit : the errors of socialism*, reprint éd., n° ed. by W. W. Bartley. ; Vol. 1 dans *The collected works of Friedrich August Hayek*, Univ. of Chicago Press [u.a.], Chicago, Ill, ISBN 978-0-226-32066-3 978-0-226-32068-7.
- Watson, J. D., F. H. Crick et others. 1953, «Molecular structure of nucleic acids», *Nature*, vol. 171, n° 4356, p. 737–738. 11186.
- Weibull, J. W. 1995, *Evolutionary game theory*, MIT Press, Cambridge, Mass, ISBN 978-0-262-23181-7.
- Wilensky, U. 1997, «NetLogo Segregation Model (Evanston, Ill. : Center for Connected Learning and Computer-based Modeling, Northwestern University)», 00004.
- Wilensky, U. 1999, «NetLogo», 02080.
- Wolfram, S. 1985, «Undecidability and intractability in theoretical physics», *Physical Review Letters*, vol. 54, n° 8, doi : 10.1103/PhysRevLett.54.735, p. 735–738, ISSN 0031-9007. 00243.
- Wolfram, S. 2002, *A new kind of science*, Wolfram Media, Champaign, IL, ISBN 978-1-57955-008-0. 05671.



## **Deuxième partie**

### **Activités de transfert et de valorisation**

---

## 6 Valorisation et transfert technologique

---

Je suis régulièrement sollicité pour appliquer mes recherches dans le domaine de l'analyse des grandes bases de données textuelles et des réseaux complexes. Ceci a donné lieu à de nombreuses collaborations de recherche entre mon laboratoire et des institutions internationales comme les Nations-Unies ou la Commission Européenne, nationales comme l'ADEME ou le MESR, ou avec des industriels et organismes privés tels qu'AXA ou l'Institut Pasteur).

Au cours de ces collaborations, j'ai développé des applications pour l'analyse et la gestion des connaissances à partir de bases de données textuelles non structurées. Celles-ci mettent l'accent sur la visualisation, l'interactivité et la prise en compte des aspects temporels de la production des connaissances.

Tous ces outils et méthodes ont été développés sous des licences libres (GNU GPLv3), garantissant une utilisation dans un cadre plus large, notamment académique. Le caractère libre et *open source* des logiciels développés a toujours été une condition préalable à de telles collaborations engendrant bien souvent des négociations supplémentaires avec mes partenaires pour leur faire comprendre l'avantage de cette approche. Cette position de principe, dont le nombre de contrats reflète la pertinence, a permis de mieux faire connaître les possibilités du libre auprès des industriels tout en créant de la valeur pour le tissu de la recherche.

### 6.1 Outils heuristiques pour la gestion des connaissances

Chacun a déjà été confronté au désarroi causé par le déluge de références renvoyées par un moteur de recherche pour peu que la question posée soit un tant soit peu générale. Quels sont les débats du Web sur le changement climatique ? Quels sont les risques émergents ? Quels sont les controverses en cours dans le domaine de la santé ? Autant de questions où la recherche d'information classique atteint rapidement ses limites. Les moteurs de recherche sont puissants lorsque l'on sait ce que l'on recherche, mais donnent le vertige lorsqu'on leur pose des questions générales.

Les outils développés dans le cadre de mon activité de transfert technologique et de valorisation contribuent à une inversion de perspective qui tire partie des masses de données au lieu d'en faire un obstacle : en appliquant mes recherches dans le domaine de l'analyse des réseaux complexes et de la scientométrie, j'ai montré qu'on pouvait proposer un point de vue synthétique et opérationnel sur des questions relativement générales par des pistes d'exploration pré-structurées plutôt qu'une longue liste de références comme le proposent de nombreux moteurs de recherche. Ce faisant, j'ai pu fournir de puissants outils heuristiques et rendre la connaissance navigable.

Ces recherches, intrinsèquement interdisciplinaires, mobilisent des domaines aussi variés que le traitement automatique du langage, l'analyse des réseaux complexes ou la visualisation d'information.

L'ambition est de transformer en profondeur notre rapport à la connaissances et plus généralement aux traces numériques produites par nos sociétés en permettant des analyses à la volée sur de grands corpus textuels non structurés.

## Retombées sur la recherche

Cette activité de valorisation et de transfert a eu plusieurs types de retombées dans le domaine de la recherche et de l'enseignement.

Elle a permis de tester dans des situations réelles et avec de vrais utilisateurs des concepts développés dans un cadre de recherche. Ceci a permis de les améliorer et de corriger les méthodes et algorithmes associés. Il est en effet assez fréquent dans le domaine de l'analyse des masses de données d'obtenir des résultats théoriques élégants, ou des analyses de données qui exhibent des motifs intéressants, sans pour autant que cela corresponde à des usages ou des objets pertinents du monde réel. Il est très difficile de s'en apercevoir en dehors d'une utilisation experte ou massive des résultats produits.

Cette activité a permis également de maintenir depuis 2009 une équipe d'ingénieurs qui a développé des outils libres et ouverts, actuellement utilisés dans le milieu académique et dans l'enseignement supérieur. Suite au logiciel Tinasoft<sup>29</sup>, Gargantext<sup>http://gargantext.org</sup>, actuellement en développement dans le cadre de plusieurs collaborations de recherche, est utilisé autant par des chercheurs que pour l'enseignement au sein d'organismes tels que les Mines ParisTech ou Sciences-Po.

La démocratisation de ces outils qui relèvent des macroscopes de Rosnay (1975) fait partie d'un engagement éthique substantiel à leur développement (cf. Chavalarias (2016)).

## 6.2 Logiciel d'analyse et de cartographie de corpus non structurés

Mes travaux sur la cartographie et la phylométrie des sciences [Chavalarias et Cointet (2008, 2013a); Cointet et Chavalarias (2008)] ont été produits grâce à WordsEvolution, un programme fait sous Matlab dont je suis le principal auteur (>95% du code). Ce programme n'étant pas très portable, j'ai entrepris depuis 2009 de faire de la valorisation et du transfert de mes résultats de recherche vers les sphères académiques, citoyenne et économiques via le développement de logiciels libres et ouverts<sup>30</sup>. J'ai ainsi piloté plusieurs projets logiciels visant à proposer des plate-formes tout-en-un pour l'analyse des corpus textuels non structurés :

- **Tinasoft**. Tinasoft [Chavalarias et collab. (2011b)] a été développé dans le cadre du projet TINA (EU FP7 2009-2011) dont j'ai été le porteur. C'était à ma connaissance le premier logiciel *open source* permettant d'aller de l'analyse des textes bruts vers l'exploration interactive de cartes de sciences avec retour contextuel aux documents<sup>31</sup>. Il se déclinait sous une version desktop et serveur. *Site web* : <http://iscpif.fr/chavalarias/softs/tinasoft>
- **Gargantext**. Gargantext [Delanoë et collab. (2015)] est un projet ambitieux, successeur de Tinasoft qui apporte plusieurs objectifs supplémentaires dont : être entièrement développé avec des technologies Web, être collaboratif, passer à l'échelle. Ce logiciel est sorti en version *beta* en décembre 2015 sous la coordination d'Alexandre Delanoë et de moi-même. Il a déjà plus d'une centaine d'utilisateurs actifs et est

29. <http://tinasoft.eu>

30. Voir les dépôts logiciels <https://github.com/moma> et <https://github.com/davidchavalarias>

31. Ce n'est pas bien sûr le premier logiciel d'analyse de corpus, nous pouvons mentionner dans préten-tion d'exhaustivité ?, Callon et collab. (1991b) et Chateauraynaud (2003)

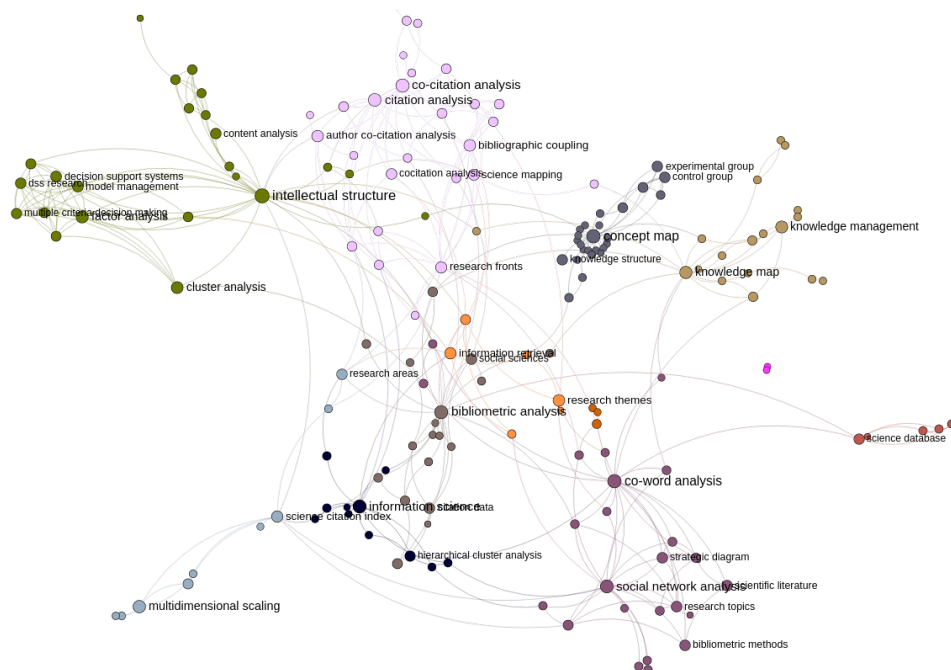


FIGURE 32 – Exemple de cartographie générée par Gargantext Cartographie du domaine de la cartographie des sciences avec Gargantext, base WoS, requête ("knowledge map" OR "science map" OR "semantic map" OR "science mapping" OR "co-word" OR "co-citation" OR "cocitation" OR "co-term" OR "science mapping" OR "mapping science" OR "mapping of science" OR "concept map" OR "mapping research").

utilisé lors de cours de controverses par plusieurs grandes écoles dont Science-Po, les Mines ParisTech et Ponts ParisTech. *Site web* : <http://gargantext.org>

Enfin, Jean-Philippe Cointet a recodé et adapté le code de WordsEvolution en lui ajoutant un important module de text-mining pour développer la plateforme Cortext (<http://cortext.net>)

## 6.3 Gestion de communautés et des grandes organisations

J'ai appliqué ces travaux dans le domaine de la cartographie de compétences et la gestion des réseaux sociaux afin de proposer des interfaces pour interagir avec des grandes organisations en tant que réseaux de collaborations et de compétences.

### 6.3.1 Community Explorer

*Tinaweb*, le module de visualisation que j'ai développé dans le cadre d'une collaboration avec la CE pour la cartographie des connaissances a été valorisé au sein de la communauté scientifique à travers une plateforme proposant de nouvelles manières d'interagir avec ses membres (<http://communityexplorer.org>).

Le *Community Explorer* (fig. 34) est un service que j'ai conçu et que je pilote. A la disposition de la communauté des systèmes complexes, il permet de cartographier à la volée les membres de n'importe quelles sous-communautés sur des critères tels que les thèmes de recherche, l'affiliation ou la nationalité.

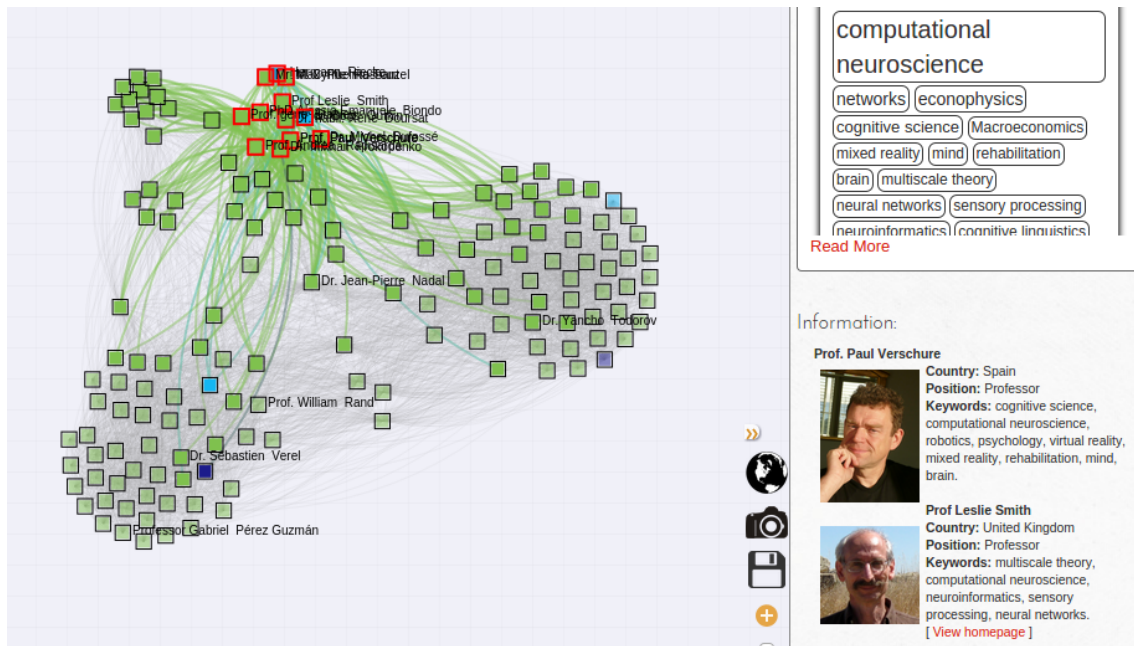


FIGURE 33 – **Le community explorer** <http://communityexplorer.org>. Le community explorer permet de naviguer dans le réseau socio-sémantique construit à partir d'un répertoire de compétences ou d'un annuaire. Il a été appliqué au répertoire de la Société des Systèmes Complexes (CSS) et comporte plus de 2300 entrées. Cette capture d'écran montre la vue du voisinage social local d'un chercheur. On distingue nettement les différentes communautés de recherche avec lequel le chercheur a des affinités et ainsi que leurs membres (ici est sélectionnée la communauté neurosciences computationnelles).

Avec plus de 2300 scientifiques utilisateurs - dont plus de 1300 à l'étranger -, cette plate-forme internationale a démontré l'intérêt d'une telle approche pour la gestion des grandes organisations, autant pour une visibilité interne qu'externe. Des collaborations de recherche sont en cours avec plusieurs services du CNRS et de l'Institut Pasteur pour la transposer à l'échelle de leur organisme.

Dans le futur, ce type de service centré sur la visualisation des réseaux multi-partites changera la manière dont les membres des grandes organisations se représentent leur environnement et interagissent avec leurs collaborateurs.

### 6.3.2 CNRS Explorer

A titre expérimental, une collaboration est en cours avec la DIRE du CNRS afin d'évaluer l'utilité d'une plate-forme comme le *community explorer* pour cartographier les unités CNRS. A partir d'une extraction *Labintel*, nous avons produit une cartographie interactive des unités CNRS et de leurs relations thématiques.

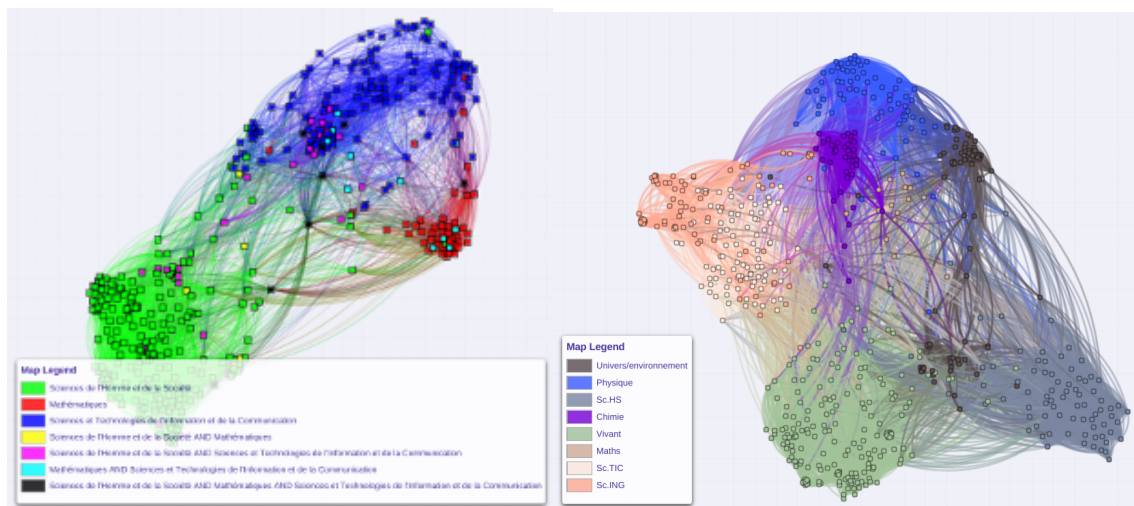


FIGURE 34 – **Explorateur d'unité CNRS.** Capture d'écran de l'explorateur d'unité du CNRS avec cartographie des unités relevant des SHS, des TIC ou des Mathématiques (environ 480 unités). La carte interactive permet de visualiser des groupes d'unités, leurs rattachements thématiques et le paysage sémantique associé. A gauche, les unités reliées par leurs disciplines, à droite, les disciplines reliées par les unités. Développeur : Samuel Castillo.

## **Troisième partie**

# **Encadrement, animation et management de la recherche**

## 7 Responsabilités de direction de laboratoire

### 7.1 Direction de l'Institut des Systèmes Complexes de Paris Île-de-France

Après avoir été directeur adjoint du GIS Institut des Systèmes Complexes de Paris Île-de-France (ISC-PIF) de 2010 à 2012, j'en suis devenu le directeur en janvier 2013 ainsi que le coordonnateur du DIM *Problématiques Transversales aux Systèmes Complexes*. En janvier 2014, j'ai pris la direction de l'ISC-PIF en tant que nouvelle unité CNRS (UPS 3611).

L'ISC-PIF est une unité atypique pour plusieurs raisons. Elle a vocation à fédérer des EPST et Universités de la Région Île-de-France autour de l'étude des systèmes complexes et elle a, à ce jour, des conventions avec 16 établissements nationaux ou franciliens qui contribuent à son fonctionnement et à sa gouvernance. Occupant 400m<sup>2</sup> dans le 13<sup>ème</sup> arrondissement de Paris, elle comporte une pépinière à projets, des espaces de séminaires ouverts à la communauté et développe des infrastructures mutualisées (cf. figures 35 & 36). Le développement de la recherche à l'ISC-PIF repose sur le principe de partenariats avec des laboratoires qui y portent des projets.

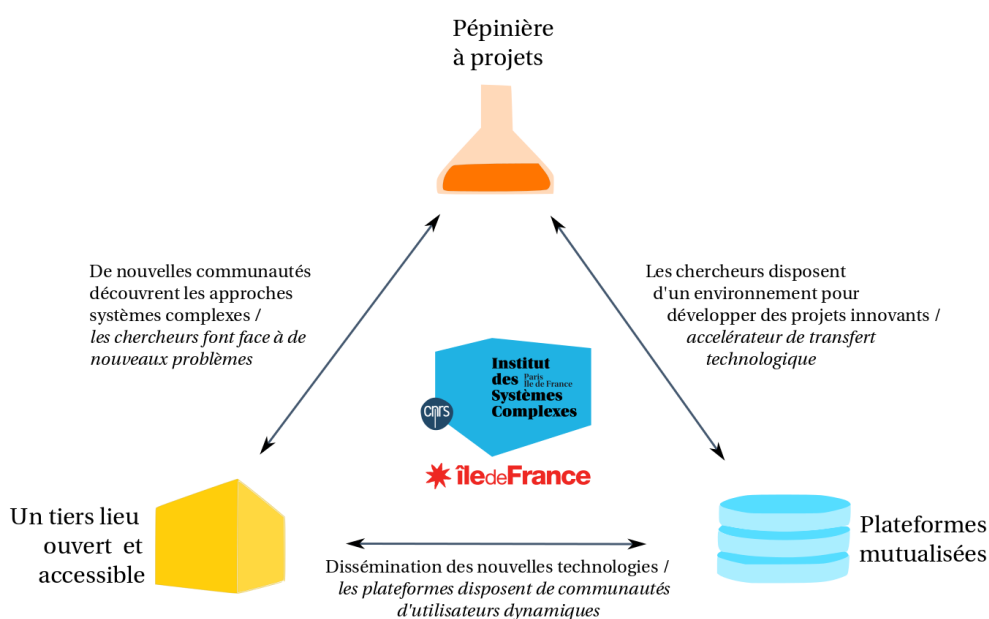


FIGURE 35 – Les trois piliers de l'action de l'ISC-PIF

Depuis 2014, le CNRS/INSHS apporte un fort soutien au développement de l'ISC-PIF à travers des moyens humains et une contribution qui constitue 20% des dépenses de fonctionnement incompressibles. Le reste est couvert par les partenaires, les projets hébergés et les projets portés par l'ISC-PIF (exemple en 2014, le budget réalisé - 1,5M€-



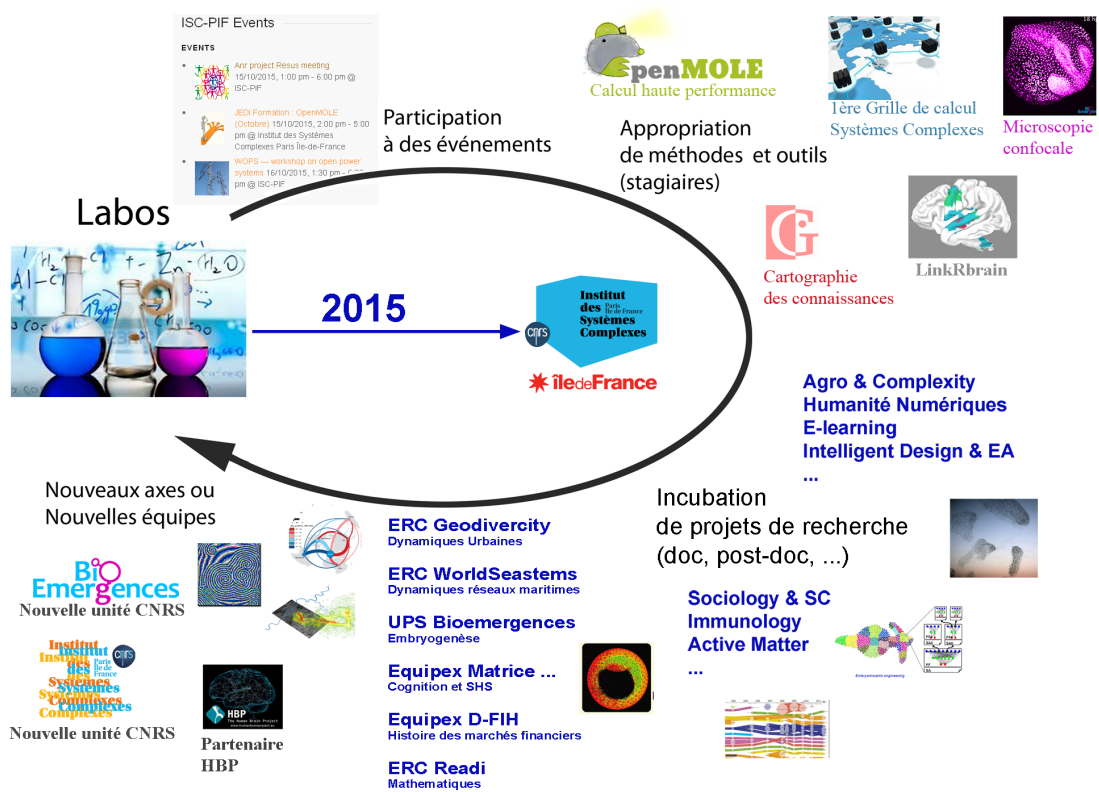


FIGURE 36 – Cycle de vie des projets à l’ISC-PIF. Panorama de l’activité en 2015 (séminaires, plateformes et projets hébergés).

se décomposait en 591k€ géré en propre, 112k€ d'apports mutualisés et 800k€ redistribués via des appels à projets).

Lorsque j'ai pris la direction de l'ISC-PIF début 2013, beaucoup pariaient sur sa fermeture fin 2013 en raison de deux facteurs. D'une part, le budget était en déficit chronique de 40k€ depuis 2011 en raison d'un changement de règles de la Région dans la gestion du DIM. D'autre part, le fond de roulement de sa trésorerie allait être épuisé. En conséquence l'Institut allait devoir rendre ses locaux à la fin 2013 et n'avait pas de solution de rechange.

Redresser l'ISC-PIF dans un contexte de crise du secteur de la recherche a nécessité d'importantes transformations auxquelles j'ai dû consacrer beaucoup de temps et d'énergie ces trois dernières années (voir section ??).

Pendant cette période, le nombre de personnes travaillant sous ma responsabilité est passé de 3 à 13 tandis que les coûts de fonctionnement incompressibles baissaient.

Par ailleurs, je suis administrativement le chef d'établissement du bâtiment de l'ISC-PIF, qui héberge quotidiennement une quarantaine de personnes plus des séminaires (au maximum 140 personnes sur site).

La direction de l'ISC-PIF s'accompagne par ailleurs d'un important travail de coordination et de justification des activités de la structure avec deux conseils d'administration par an (avec leurs rapports autour de 50 pages chacun), un conseil d'Orientation et de Surveillance par an (et son rapport), l'organisation de comités de pilotages en moyenne une fois par mois et la rédaction de rapports d'activité annuels pour la région Île-de-France.

---

## 8 Responsabilités de coordination d'un réseau régional de recherche

---

Depuis 2013, je suis le coordonnateur du DIM *Problématiques Transversales aux Systèmes Complexes*. J'ai la responsabilité directe de la bonne allocation de la subvention régionale d'une enveloppe moyenne de 1M€ par an.

Dans le cadre de cette coordination, j'ai supervisé notamment l'organisation des appels à projets d'un point de vue scientifique et opérationnel, en collaboration avec le comité de pilotage de l'ISC-PIF. Cet appel à projet a eu beaucoup de succès et a reçu chaque année (2011 à 2015) entre 120 et 150 dossiers, toutes catégories confondues (chacun étant évalué par deux évaluateurs hors Île-de-France).

J'ai également défini et mis en œuvre plusieurs grandes opérations d'investissement dans le cadre du DIM et de l'ISC-PIF.

1. **Maison des Systèmes Complexes** Mise en place d'un partenariat avec la Ville de Paris, la Région et le CNRS pour la rénovation des locaux de l'ISC-PIF au 113 rue Nationale, Paris 13. Montant total de l'opération : 558k€. Étapes : négociations avec la Ville de Paris pour une proposition de locaux à coûts modérés (2 ans) ; négociations avec la région pour accepter le montage en co-financement avec la Ville de Paris (6 mois), négociations avec le CNRS pour accepter le montage proposé (6 mois), tâches de maîtrise d'ouvrage sur les locaux (6 mois : définition des besoins, suivi des travaux), mise en place de la vie sur site (18 mois : déménagement, règlements, etc.).

2. **Renouvellement de la grille de calcul de l'ISC-PIF** Renouvellement de l'infrastructure de grille de calcul associée à la V.O. européenne "Complex Systems" en partenariat avec le LAL et le LPNHE dans le cadre de GRIF (Grille au service de la Recherche en Ile-de-France). Montant total de l'opération : 600k€.
3. **Renouvellement du parc informatique de l'ISC-PIF** Renouvellement des serveurs de l'ISC-PIF en partenariat avec les projets hébergés : 110k€.

Toutes ces opérations se sont accompagnées d'un important travail de négociation en amont et de justification en aval. J'ai bénéficié pour cela des aides précieuses de la Secrétaire Générale Laurence Gareaux de l'ISC-PIF et de la responsable partenariats et valorisation Catherine Robert.

---

## 9 Responsabilités dans l'animation de programmes et projets de recherche

---

En tant que chercheur au CAMS et directeur de l'ISC-PIF, j'anime une équipe de recherche sur les humanités numériques et les sciences sociales computationnelle qui relève des deux laboratoires.

Dans ce cadre, je participe à de nombreux projets nationaux ou internationaux (cf. CV) dont je suis pour certains le responsable scientifique et le porteur (pour les contrats hors agences de financement de la recherche, voir la section 6, pour les projets passés, voir mon CV) :

- **2013-2017 Science en Poche** : Porteur du projet, Crowdsourcing et expériences scientifiques utilisant les terminaux mobiles, Programme Émergence(s) de la ville de Paris.
- **2012-2016 FORCCAST** : Responsable scientifique pour le CAMS et l'ISC-PIF, analyse de controverses, Projet Investissement d'Avenir IDEFI coordonné par Sciences-Po.
- **2015-2019 Algodiv** : Responsable scientifique pour le CAMS, Recommandation algorithmique et diversité des informations du web, Projet ANR coordonné par le centre Marc Bloch, CNRS.
- **2016-2020 EPIQUE** : Responsable scientifique pour l'ISC-PIF, "Reconstruction à grande échelle des réseaux phylomemetiques", Coordonné par le LIP6, UPMC.

---

## 10 Autres Responsabilités

---

J'ai participé à diverses instances nationales ou internationales en tant que membre nommé ou membre élu :

- Participation à la commission régional d'inter-classement de la BAP E en 2015,
- Expert indépendant pour la commission européenne 7ème Programme cadre - Future Emergent Technologies.
- Vice-Président Secrétaire de la Complex Systems Society (2009-2014) : gestion des membres, gestion du systèmes d'information, aide à l'organisation de la conférence annuelle, lettre de diffusion. Plus de mille membres à travers le monde.
- Membre du jury en tant que représentant du laboratoire pour les concours ITA 2015 numéros 100, 95 et 39.
- Évaluateur pour des revues et colloques internationaux (PNAS, JASSS, WCSS, MARAMI, etc.),
- Évaluateur extérieur pour l'axe "Numérique : Grande échelle & Complexité" de l'École des Mines ParisTech (avril 2014).
- Membre du groupe de réflexion de l'INSHS pour les axes "Crowdsourcing" et "sciences du comportement".