



© CNRS / Délégation PMA / Ebdelle Morel

Édito

de François-Joseph Ruggiu,
Directeur de l'InSHS

Que 2021 soit une meilleure année que 2020 : chacun l'espère pour toutes et tous et pour le monde de l'enseignement supérieur et de la recherche. Les conséquences de la crise sanitaire, sur fond d'une crise environnementale globale, pèsent, en effet, sur les vies des individus et des établissements [p2]

VIE DES RÉSEAUX

Justice et intérêt - Aujourd'hui

Questionner les sciences économiques, leur pertinence, leurs limites, leur nature scientifique même, est une entreprise sans doute aussi ancienne qu'elles-mêmes, mais dont l'urgence est rappelée à chaque crise — et c'est d'actualité [p11]

À L'HORIZON

Une biographie collective au xx^e siècle, entre histoire des migrations et histoire de la Shoah

L'objet de ce projet est d'interroger les liens entre dynamiques migratoires et persécution. L'enjeu consiste à étudier à la fois ceux qui restent et ceux qui émigrent dans différents espaces tout en interrogeant les connexions entre ces deux groupes [p14]

ZOOM SUR...

Linguistique Computationnelle

La linguistique computationnelle, comme plus généralement la linguistique, cherche à comprendre le fonctionnement du langage humain. Mais sa particularité est d'aborder cette question dans une perspective de calcul informatisé, en développant des algorithmes qui peuvent être exécutés par un programme informatique [p17]

UN CARNET À LA UNE

Littérature et Révolution

Le carnet Littérature et Révolution est une publication d'Olivier Ritz, maître de conférences en littérature française à l'université de Paris. Depuis 2016, il y interroge le traitement de la Révolution française par la littérature et le théâtre [p35]

À PROPOS

Mémoires des Archéologues et des Sites Archéologiques, le consortium MASA au service de la communauté archéologique

MASA est l'un des consortiums de la TGR Huma-Num, animé par dix partenaires interinstitutionnels parmi lesquels les Maisons des Sciences de l'Homme (MSH) jouent un rôle fondamental. Il est né du besoin impérieux de sauvegarder les archives des archéologues et des sites archéologiques [p3]

FOCUS

Quel contrôle de l'espace numérique en Russie ? Entre contraintes et contournements

Au printemps 2018, le Parquet russe demande aux concepteurs de la messagerie chiffrée Telegram de lui livrer les clés de chiffrement de leur application. Pavel Dourov, le fondateur de la plateforme, refuse, depuis l'étranger, de céder aux demandes officielles [p6]

OUTILS DE LA RECHERCHE

DicoPolHiS, le pari d'une vulgarisation exigeante de l'histoire de la santé

La première notice de DicoPolHiS a été publiée le 16 mars 2020, soit la veille du premier confinement [p8]

LIVRE

REVUE

PHOTO

À LA UNE



Le cri de Gaïa. Penser la Terre avec Bruno Latour, Frédérique Aït-Touati, Emanuele Coccia, 2021

Bruno Latour inspire depuis plusieurs décennies le travail de philosophes, historiennes et historiens,

sociologues, éthologues, anthropologues et artistes dans le monde entier. *Face à Gaïa* appelle à une réaction au-delà de la simple assimilation théorique. Ce qui a l'apparence d'une question purement scientifique est en vérité un différend d'ordre politique [...]

[voir toutes les publications](#)



La *Revue des Sciences Sociales* est l'une des plus anciennes revues françaises de sciences sociales. Fondée par Julien Freund en 1972 à Strasbourg, car-

refour de cultures, savoirs et mémoires, terre de frontière marquée par les conflits et les recompositions, elle s'est rapidement ouverte aux échanges franco-allemands et, au fil du temps, européens. Revue semestrielle à comité de lecture, elle convoque différentes disciplines autour d'un thème de société [...]

[voir toutes les revues](#)



Détail d'une tapisserie du 19^e siècle à la Cité internationale d'Aubusson

© Olivier RESNOLIN / IRAMAL CRIPA / CNRS Photothèque

Édito

de François-Joseph Ruggiu
Directeur de l'InSHS



© CNRS / Délégation RMA / Flicke Mehl

Que 2021 soit une meilleure année que 2020 : chacun l'espère pour toutes et tous et pour le monde de l'enseignement supérieur et de la recherche. Les

conséquences de la crise sanitaire, sur fond d'une crise environnementale globale, pèsent, en effet, sur les vies des individus et des établissements. Les laboratoires de sciences humaines et de sciences sociales du CNRS sont atteints en profondeur même s'ils ont su mettre en place au printemps et à l'automne derniers les actions indispensables pour la poursuite de leurs activités de recherche.

Il est important de remercier ici, à nouveau, les directeurs et les directrices de laboratoires, ainsi que tous les personnels, qui ont fait beaucoup d'efforts pour que l'année 2020 puisse se dérouler le plus normalement possible. Un énorme travail a, en particulier, été accompli pour que les crédits affectés à certaines dépenses entravées par la crise, comme les missions ou les manifestations scientifiques, puissent être reportés vers d'autres postes, comme l'équipement ou les infrastructures. Mais la période reste très compliquée pour les étudiantes et étudiants, du premier cycle au doctorat, dont l'accès aux cours et aux bibliothèques est problématique, et pour nombre de collègues confrontés, pour la plupart, à une augmentation inédite de leur charge de travail et à des difficultés accrues d'accès à la documentation indispensable à leur production scientifique ou à leurs terrains de recherche.

Dans ce contexte sombre, une touche d'optimisme : le budget de l'InSHS du CNRS en 2021 est le meilleur que nous ayons obtenu ces dernières années. Il passe, en effet, la barre des 20 millions d'euros dont 76 % (15,5 millions sur 20,3) sont consacrés aux dotations globales des laboratoires. Ces crédits leur ont été notifiés dès le mois de janvier pour qu'ils financent tout au long de l'année, et selon les modalités qui leurs sont propres, leurs actions scientifiques ou leurs frais d'infrastructures. Ces dotations constituent le soutien de base aux unités de recherche dont le CNRS est le pourvoyeur avec ses partenaires académiques. Le budget des deux Très Grandes Infrastructures de Recherche que sont [Huma-Num](#), qui offre des services numériques aux unités et aux chercheurs et chercheuses, et [Progedo](#), qui met à disposition des données massives aux spécialistes de sciences sociales, représente 2,25 millions (11 %). L'essentiel des 13 % restants va au financement des réseaux nationaux (Groupements d'intérêt scientifiques ; Groupements de recherche ; Réseaux thématiques pluridisciplinaires) ou internationaux (*International Research Programs ; International Research Networks*) que soutient l'InSHS du CNRS ou encore à des engagements pluriannuels comme le soutien aux revues ou à des projets retenus dans le cadre des Contrats de plan État-Région.

Avec les Universités, les grandes écoles et les autres organismes, cotutelles des unités, le CNRS leur donne ainsi les moyens de leur fonctionnement régulier, parallèlement aux différentes agences de financement, locales, nationales ou européennes vers lesquelles se tournent les collègues qui ont besoin de financements particuliers, individuels ou collectifs, en raison de l'ampleur ou de la nature de leurs recherches. Il est à espérer que cette hausse des crédits obtenues en 2021 se poursuivra dans les années à venir car il est clair que les besoins en sciences humaines et en sciences sociales restent très importants.

Mémoires des Archéologues et des Sites Archéologiques, le consortium MASA au service de la communauté archéologique

MASA est l'un des consortiums de la Très grande infrastructure de recherche (TGR) Huma-Num, animé par dix partenaires interinstitutionnels parmi lesquels les Maisons des Sciences de l'Homme (MSH) jouent un rôle fondamental. Il est né du besoin impérieux de sauvegarder les archives des archéologues et des sites archéologiques. Dans une première phase (2013-2016), les compétences des partenaires du consortium ont été mobilisées pour le traitement — de la numérisation à la publication — de corpus d'archives et de données archéologiques. Des preuves de concept, établies à partir de ces terrains expérimentaux, sont destinées à faire la démonstration du partage et de l'interopérabilité des données. La deuxième phase qui vient de s'achever (2017-2020) a consisté à réunir ces briques pour constituer un écosystème numérique selon le cycle de vie des données, conforme aux principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable). L'objectif de la troisième étape qui s'ouvre maintenant est la diffusion des outils et des bonnes pratiques auprès de la communauté archéologique.

Le consortium est fondé sur des constats en grande partie partagés avec d'autres disciplines des sciences humaines et sociales. Tout d'abord, celui inhérent à la discipline elle-même : la fouille archéologique est, par définition, en grande partie destructrice de son propre objet d'étude. L'expérience n'étant pas renouvelable, l'enregistrement des données sur le terrain constitue, avec l'ensemble du matériel recueilli, des archives de fouilles irremplaçables, qualifiables de données primaires. Le deuxième constat porte sur la constitution de cette documentation qui donne lieu à des fonds d'archives identifiables à partir d'une personne ou d'un site. Ensuite, vient la nature des cahiers de fouilles, construits sous des formes plus ou moins structurées, du journal de bord relatant le récit des découvertes au système d'information selon des unités d'enregistrement formalisées. Finalement, le constat principal est celui d'une formidable hétérogénéité, elle-même porteuse d'informations. Elle concerne aussi bien la nature des archives (textes, photographies, plans, relevés, croquis, mesures...), que leurs formats (papier ou numérique), ou encore la structure de l'information (récit, systèmes d'information). Les étapes successives de l'informatisation, rythmée par l'accessibilité à de nouveaux outils techniques et/ou méthodologiques, ont chacune apporté leur lot de tentatives d'homogénéisation universelle, la plupart avortées ou fort heureusement abandonnées. Ces expériences ont elles-mêmes produit de nouvelles archives nativement numériques (en particulier dans les années 1990-2000) particulièrement hétérogènes par leurs formats, leurs structures, leurs supports. Afin de répondre aux besoins de sauvegarde et de mise à disposition des fonds d'archives, ainsi que d'interconnexions et d'interrogations croisées des jeux de données, les efforts du consortium MASA ont été déployés selon deux axes. Le premier porte sur les bonnes pratiques d'archivage et les plans de gestion de données, qui sont de plus en plus considérés par les agences de financement comme des livrables

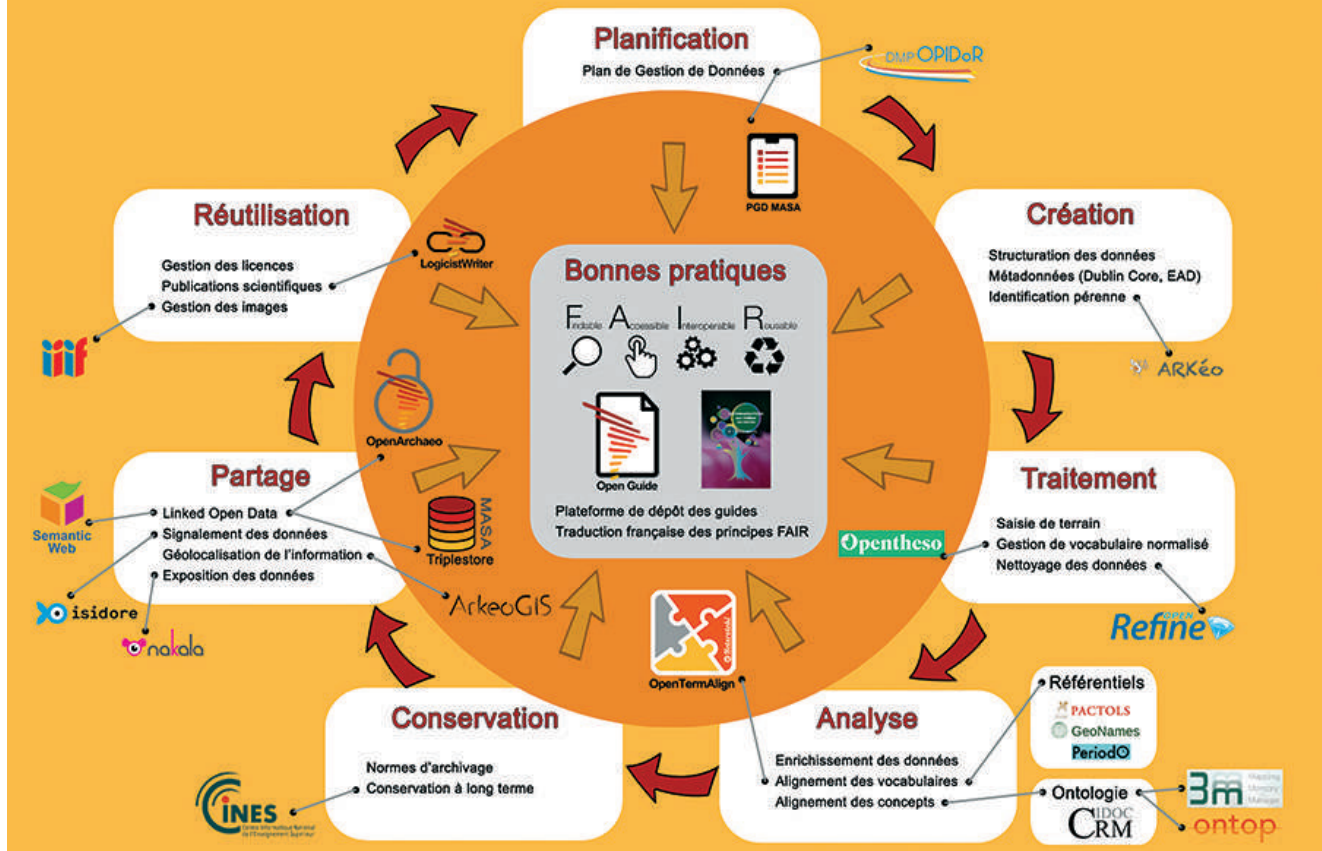


Carte des partenaires du consortium MASA

obligatoires des projets de recherche. Les travaux du second sur l'interopérabilité ont conduit à s'orienter vers le web sémantique.

Résolument ancré dans la dynamique de la science ouverte, MASA est un vecteur de diffusion des principes FAIR dans la communauté archéologique. L'objectif est de favoriser la FAIRisation des données, à l'image de celles mises à disposition par les partenaires du consortium, qui alimentent progressivement la plateforme OpenArchaeo dédiée à la publication de jeux de données dans le web sémantique. Le consortium bénéficie des compétences exceptionnelles de chercheurs, chercheurs et ingénieur(e)s des universités et du CNRS qui, chacun dans leur domaine, ont permis d'atteindre collectivement un très haut niveau d'expertise sur chaque étape de la conduite d'un projet numérique. Cette base robuste et l'engagement déterminé du consortium dans l'application et la diffusion des principes FAIR a conduit à constituer un écosystème numérique structuré selon le cycle de vie des données. Il mobilise des méthodes, des référentiels, des standards et des outils existant pour répondre aux besoins des projets mis en œuvre. L'expertise des partenaires du consortium et les expériences conduites sur différents corpus ont permis d'identifier les méthodes et outils à réutiliser, celles et ceux à consolider et, enfin, les manques à combler. L'écosystème MASA fait ainsi appel à des outils *open source* (OPIDoR, ARKéo, OpenRefine, 3M, Ontop, Nakala, Isidore, IIF, GeoNames, PerioO, CIDOC CRM...). Il soutient fortement le développement de certains (Opentheso¹, PACTOLS², ArkeoGIS). Enfin, il réalise des développements appropriés (Plan de Gestion de Données pour l'archéologie³, OpenTermAlign, OpenGuide, OpenArchaeo,

1. Voir aussi <https://github.com/miledrousset/opentheso>.
 2. Voir aussi <https://pactols.frantiq.fr/opentheso/>.
 3. Voir aussi https://dmp.opidor.fr/public_templates, rechercher « MASA ».



L'écosystème numérique MASA selon le cycle de vie des données et les principes FAIR

LogicistWriter). Dans tous les cas, le consortium veille au respect des standards nationaux et internationaux, en application des principes FAIR.

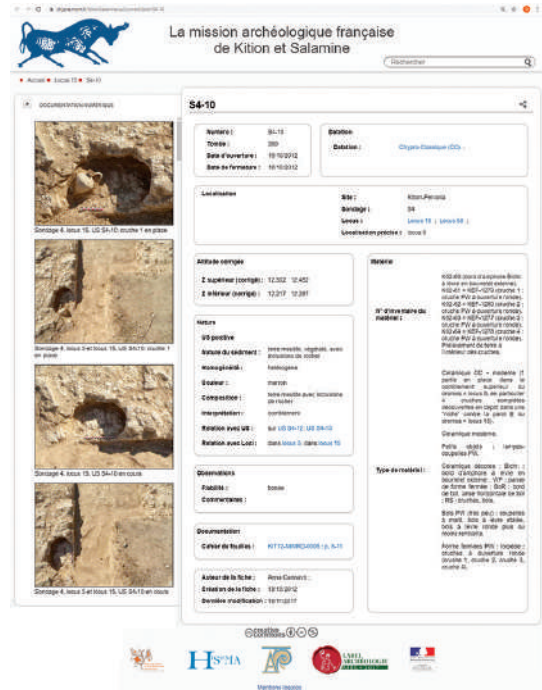
À titre d'exemple, le consortium a fait le choix du *CIDOC Conceptual Reference Model (CRM)*, l'ontologie⁴ de domaine dédiée au patrimoine culturel, coordonné par un consortium international (CRM SIG). Le rôle du CIDOC CRM, pour MASA, est de servir de surcouche unique et transversale pour faire communiquer des jeux de données hétérogènes publiés sur le web, afin de les rendre interopérables sans modifier ni le format ni la structure des bases de données initiales.

MASA a développé, en partenariat avec la société SPARNA, spécialisée en Web sémantique, et le Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT), la plateforme web sémantique *OpenArcheo*. Exploitant les *triplestores*⁵ archéologiques appariés avec le CIDOC de manière fédérée, OpenArcheo a été conçu avec deux interfaces distinctes. La première est un classique *Endpoint SPARQL* pour des requêtes en langage SPARQL (langage dérivé du SQL pour interroger le web sémantique) couplé à des services web pour exploiter le *triplestore* MASA. Cette première interface, dont la vocation est d'être exploitée par des machines, est davantage accessible à des informaticiens. La *seconde*, résolument plus conviviale, est basée sur un système d'icônes et de concepts bien maîtrisés par

les archéologues (Unité stratigraphique, Mobilier, Fait, Site...). OpenArcheo a fait l'objet du développement du composant *SPARNATURAL* (pour produire une requête SPARQL à partir du langage naturel), modulable et réutilisable pour lequel les manifestations d'intérêt se multiplient dans d'autres domaines que l'archéologie.

Les travaux de MASA ont un écho au-delà des frontières qui se traduit par des collaborations dans des projets internationaux tel que le programme Horizon 2020 de l'Union européenne *ARIADNEplus*, qui rassemble plus de quarante partenaires en archéologie et en informatique pour offrir une infrastructure de partage des données, ou le COST *SEADDA* qui porte sur l'avenir des données archéologiques en Europe et au-delà. MASA travaille non seulement avec d'autres consortiums d'Huma-Num (*3D-SHS* et *CAHIER*, par exemple), mais aussi d'autres communautés dans le champ du patrimoine culturel qui sollicitent MASA sur les enjeux de l'interopérabilité et du web sémantique. C'est, par exemple, le cas du groupe de travail *Données numériques du Chantier Notre-Dame* (CNRS/Ministère de la Culture), de celui sur *SPARNATURAL* avec les Archives nationales et la Bibliothèque nationale de France, ou encore avec le *projet européen 4CH*, *Competence Centre for the Conservation of Cultural Heritage*. Cette double ouverture, interdisciplinaire, vers le patrimoine culturel d'une part et l'intelligence artificielle d'autre part, est une reconnaissance des travaux et de l'expertise du consortium. Elle invite à

4. Description des concepts d'un domaine et leurs relations.
5. Entrepôt de données structurées en triplets Sujet-Prédicat-Objet.



À gauche, l'explorateur de la plateforme OpenArcheo donnant accès aux bases de données initiales ; ici, les fouilles du site de Kition-Pervolia (Chypre) à droite

poursuivre ces collaborations essentielles et l'investissement en recherche, en particulier sur le web sémantique, la fouille de données et la réutilisation des données (le R des FAIR) qui sont des enjeux majeurs de la science ouverte. C'est ce qui permettra l'ouverture vers de nouveaux champs de recherche fondés sur l'exploitation de grands corpus archéologiques publiés dans le web des données, soit en réunissant virtuellement des fonds dispersés, soit en faisant émerger des questionnements inédits lorsqu'il sera possible d'exploiter les règles d'inférences dans les graphes de données.

et Unités mixtes des Instituts français de recherche à l'étranger - UMIFRE —, universités, CNRS et sociétés savantes). Cette rencontre permettra de présenter l'écosystème numérique et la palette d'outils qui le compose, ainsi qu'une offre de formation aux bonnes pratiques à travers les principaux outils (MASA a, par exemple, mis en place des formations au CIDOC CRM à l'aide d'un jeu de cartes). L'objectif de cette rencontre est de diffuser largement les pratiques et les outils pour la mise à disposition de corpus archéologiques interopérables dans des plateformes nationales et internationales.

En 2021, MASA organisera une rencontre nationale réunissant la communauté archéologique française dans toutes ses dimensions (ministère de la Culture, Conseil national de la recherche archéologique, archéologie préventive publique — Inrap et collectivités territoriales — et privée, archéologie française à l'étranger — Réseau des Écoles françaises à l'étranger

Laurence Rageot, MSH Val de Loire, responsable administrative du consortium MASA ; Xavier Rodier, MSH Val de Loire, coordinateur scientifique du consortium MASA

► S'abonner à la liste d'information

contact&info

- Laurence Rageot, MSH VL
laurence.rageot@univ-tours.fr
- Xavier Rodier, MSH VL
xavier.rodier@univ-tours.fr
- Pour en savoir plus
<https://masa.hypotheses.org>



Quel contrôle de l'espace numérique en Russie ? Entre contraintes et contournements



Les « clés » de Telegram ironiquement transmises par Pavel Durov au FSB, 10 avril 2018 © Agora

Au printemps 2018, le Parquet russe demande aux concepteurs de la messagerie chiffrée Telegram de lui livrer les clés de chiffrement de leur application. Pavel Durov, le fondateur de la plateforme, refuse, depuis l'étranger, de céder aux demandes officielles. En réponse, l'agence Roskomnadzor, chargée de la régulation de l'internet russe, ordonne le blocage de l'application sur le territoire national, ce qui suscite des manifestations à Moscou et conduit Telegram à mobiliser la communauté technique et civile pour déployer une stratégie de contournement du blocage qui le rend inopérant. Entre 2018 et 2020, Telegram est officiellement bloqué en Russie mais ses nombreux usagers continuent à l'utiliser, et sa popularité ne cesse de croître. Cet exemple de contournement est l'un des plus éclairants dans l'histoire récente d'internet en Russie. Il met en lumière non seulement les ambitions régulatrices des instances de contrôle officielles mais aussi l'aspect protéiforme des mobilisations qui s'y opposent, réunissant des acteurs aux aspirations politiques et économiques variées, sensibilisant de larges couches de la population au contournement et à la réflexion sur les droits numériques. Cet exemple ne peut cependant, à lui seul, rendre compte de la richesse des pratiques et des usages, des plus sophistiqués aux plus modestes, des utilisateurs russes qui s'adaptent aux diverses contraintes en ligne.

L'équipe du projet ANR ResistTIC « Les résistants du net. Critique et évasion face à la coercition numérique en Russie » s'est ainsi lancée, à partir de 2018, dans l'exploration des contraintes numériques et de leurs contournements sur le terrain russe. Au fil d'entretiens et d'observations, en ligne et hors ligne, les données collectées ont permis de réaliser une sociologie des savoirs résistants en Russie et d'écrire une histoire controversée de l'internet russe¹. Cette dernière s'inscrit parfaitement dans le mouvement de déception à l'égard des promesses d'internet qui, d'outil de la démocratisation et du partage, serait devenu un instrument majeur de la surveillance et du contrôle des populations. Ce tournant mérite examen pour mieux comprendre

comment s'élaborent non seulement les nouvelles emprises sur l'espace numérique mais aussi les stratégies qui viennent contrarier les dispositifs d'encadrement. En la matière, le cas russe est particulièrement fécond pour nourrir les savoirs empiriques et la réflexion scientifique sur les alternatives en ligne.

Comprendre les enjeux actuels du numérique en Russie nécessite d'abord de saisir le caractère spécifique de l'écosystème local, et comment il s'inscrit dans la réalité globale. Les débuts d'internet en Russie, dans les années 1990, ont coïncidé avec les réformes libérales qui ont suivi la disparition de l'Union soviétique. Dans un contexte de brutale ouverture au monde et au marché, les premières expérimentations se sont déroulées à toutes les échelles, des plus locales aux plus globales, donnant naissance à un environnement numérique à la fois connecté aux évolutions internationales et spécifique par la multiplicité de ses acteurs, l'apparition de champions nationaux et, jusqu'à la fin des années 2000, la faiblesse relative de la régulation publique. L'État russe a soutenu le déploiement des infrastructures numériques sur l'ensemble du territoire, permettant à la majorité de la population de se connecter au réseau et d'accéder à la fois aux services des plateformes internationales (GAFAM) mais aussi à de puissants acteurs privés locaux comme Yandex (le moteur de recherche russe concurrent de Google) ou VKontakte (le réseau social concurrent de Facebook). Dans cet espace foisonnant, servi par une multitude de fournisseurs d'accès à internet, les citoyens ont pu découvrir des contenus éclectiques, allant des médias alternatifs aux sites d'e-commerce, en passant par les blogs citoyens, les bibliothèques pirates en ligne ou la musique à la demande. Ce dynamisme numérique a porté des espoirs de « modernisation économique » et de « transition démocratique », pensés dans des termes proches des réformes du début des années 1990, espoirs qui ont culminé lors des manifestations contre les fraudes électorales en 2011-2012. Les réseaux sociaux et les médias en ligne ont à l'époque facilité la coordination et la mise en visibilité du mouvement protestataire.

1. Elle permettra la création d'une *Timeline* accessible au plus grand nombre pour connaître et réfléchir aux apports du cas numérique russe.

Cette secousse démocratique a été suivie d'une réaction politique et institutionnelle à la fois claire dans ses intentions régulatrices et complexe dans sa mise en œuvre. La législation s'est considérablement renforcée dans une logique de contrôle national de l'information en ligne, des données des utilisateurs, de régulation des contenus illicites et, *in fine*, de « souverainisation ». Sa mise en œuvre a été confiée à l'agence Roskomnadzor, en coopération avec les organes de maintien de l'ordre, obligeant les fournisseurs d'accès internet à filtrer leur trafic et à donner accès à son contenu aux organes de sécurité (par l'intermédiaire des boîtiers dits SORM). Roskomnadzor a également pour mission de bloquer l'accès aux ressources dites « illégales » dont les redéfinitions permanentes constituent un enjeu de première importance aussi bien pour les pouvoirs publics que pour divers acteurs de la société civile. Le contrôle de l'espace public numérique est aussi distribué à des acteurs privés (fournisseurs d'accès à internet, plateformes locales comme Yandex ou réseau social VKontakte) qui traduisent et relaient les exigences publiques tout en défendant leurs intérêts économiques propres ; ou encore à des organisations citoyennes pratiquant le vigilantisme en ligne se donnant pour but de « purifier » le cyberspace russophone.

Mais les ambitions régulatrices, qu'elles soient portées par des acteurs publics ou privés, experts ou amateurs, se heurtent à des contraintes techniques et à l'inventivité des professionnels d'internet et de l'espace public, des militants et des entrepreneurs, ou des simples citoyens éprouvés par les barrières numériques qu'ils expérimentent dans leurs activités quotidiennes. Ainsi, les protestations sociales ou environnementales (à l'exemple de la mobilisation contre la décharge de Shies dans le nord de la Russie) montrent les usages numériques inventifs et perspicaces des citoyens mobilisés. Leurs contournements peuvent s'appuyer sur des savoirs techniques locaux ou globaux, sophistiqués ou rudimentaires. Les acteurs disséminent et banalisent des savoir-faire visant à déjouer les interdits : grâce à eux, les usagers apprennent l'anonymisation, l'usage de VPN, de serveurs proxy, du réseau TOR et d'autres pratiques de contournement qui nourrissent des savoirs ordinaires et circulent entre les différents pays post-soviétiques grâce à des réseaux professionnels, partisans ou amicaux.

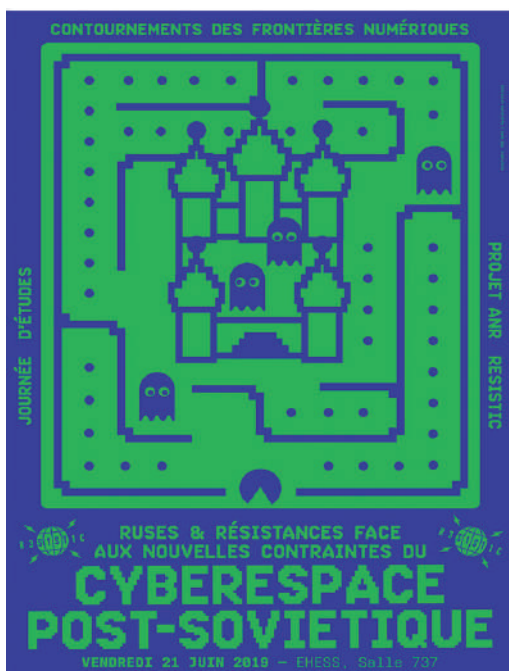


« Ne touche pas à internet ! », Manifestation contre le blocage d'internet, Moscou, 13 mai 2018 © Françoise Dauce

Dans un espace où la suspicion est historiquement centrée sur l'État, les tentatives de régulation publiques sont dénoncées comme liberticides par les associations nouvellement créées de défense des libertés numériques (parti pirate, Roskomsvoboda, Société pour la défense de l'internet...), mais la suspicion à l'égard des emprises économiques et capitalistiques retient moins leur attention. C'est ainsi que la défense pour les libertés du net est souvent menée au nom de justifications économiques soulignant les vertus de la liberté numérique et les coûts de la régulation. Les formateurs à la sécurité numérique qui travaillent avec les organisations de défense des droits de l'homme montrent une ferme hostilité aux applications russes (VKontakte ou mail.ru) contrôlées par l'État mais une grande confiance dans les plateformes internationales (GAFAM), souvent décriées dans leurs pays d'origine.

Par-delà la dialectique opposant libertés et contrôles en ligne, l'étude des critiques et contournements numériques dans la société russe permet de documenter l'émergence de réflexivités et de solidarités nouvelles sur internet. Ces savoirs valent pour cet espace spécifique mais répondent aussi aux interrogations qui surgissent en d'autres régions du monde confrontées elles aussi à des processus, affichés ou latents, de souverainisation ou de contrôle d'internet.

L'équipe du projet ResisTIC est composée de Olga Bronnikova, Françoise Dauce, Fabrice Demarthon, Ksenia Ermoshina, Valery Kossov, Benjamin Loveluck, Francesca Musiani, Bella Ostromooukhova, Perrine Poupin, Anna Zaytseva.



Programme de la journée d'étude du projet ResisTIC du 21 juin 2019 © OKB-Buro

contact&info

- ▶ Françoise Dauce, Cercoc
francoise.dauce@ehess.fr
- Bella Ostromooukhova, EUR'ORBEM
ostrob@gmail.com
- ▶ Pour en savoir plus
<https://resistic.fr>

DicoPolHiS, le pari d'une vulgarisation exigeante de l'histoire de la santé

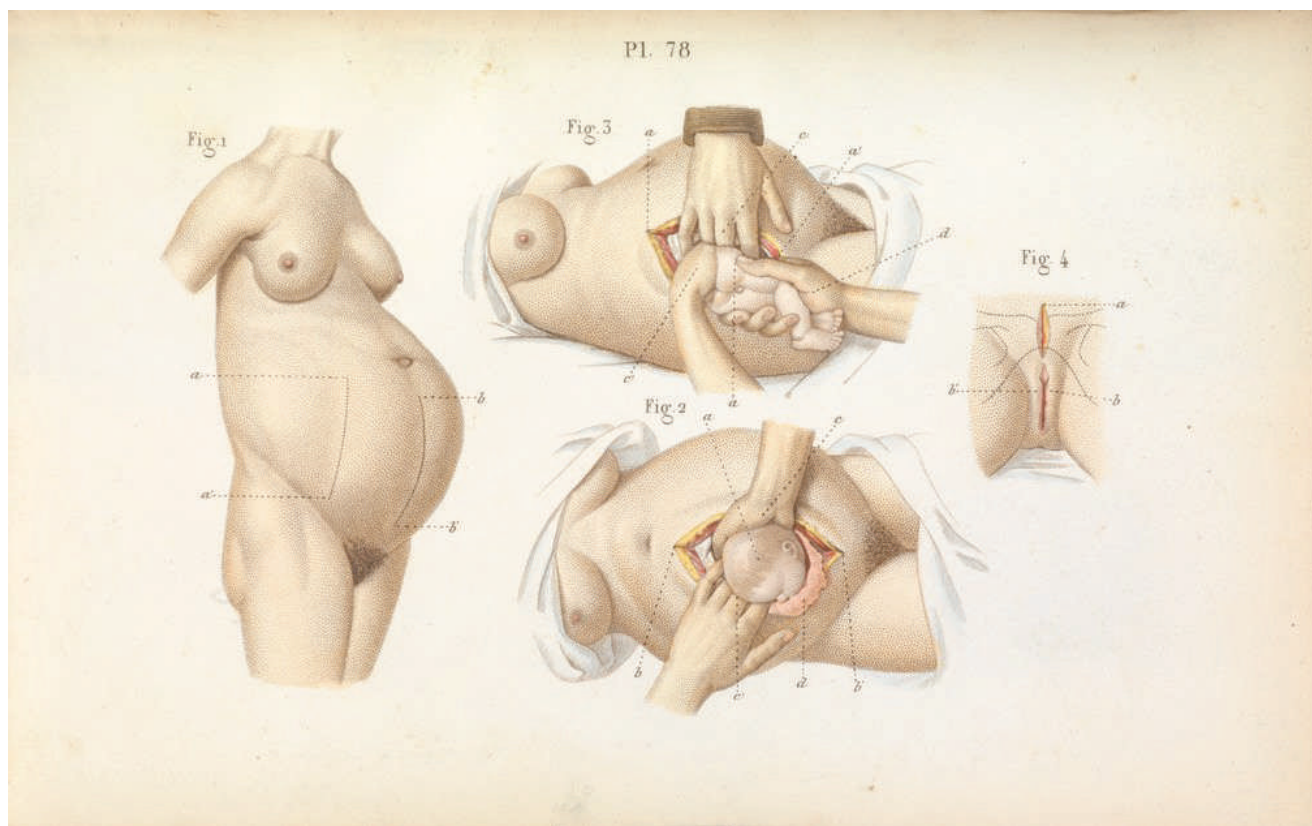
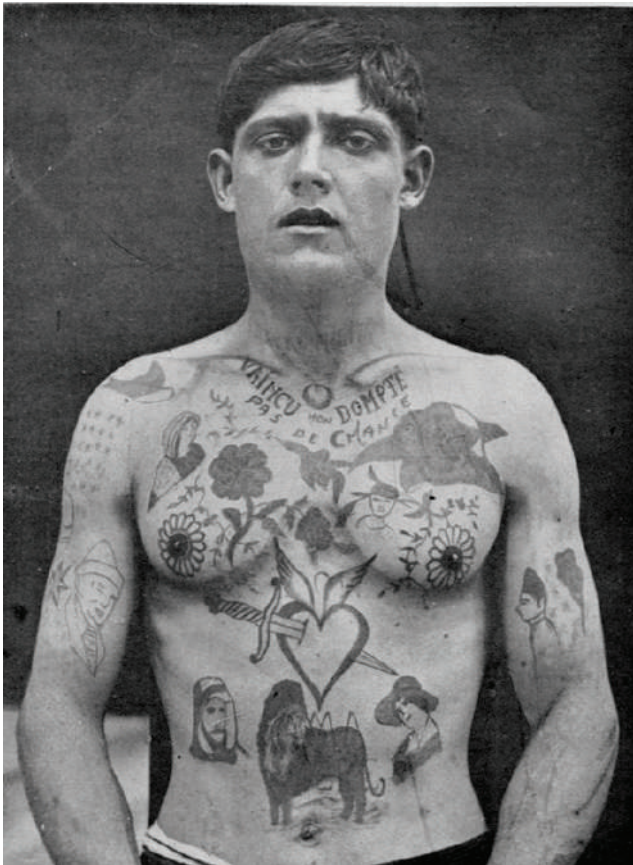


Plate 78, Surgical technique for caesarean section, Wellcome collection. Attribution 4.0 International (CC BY 4.0)

La première notice de *DicoPolHiS* a été publiée le 16 mars 2020, soit la veille du premier confinement. Drôle de baptême pour le *Dictionnaire Politique d'Histoire de la Santé* ! Mais cette coïncidence chronologique ne manque pas de sens car cette entreprise éditoriale d'un nouveau genre s'est donnée pour objectif premier de faire comprendre au grand public les enjeux politiques majeurs qui irriguent nos manières de penser la santé. À l'heure où se produisent de vifs débats concernant les vaccins, la quarantaine, les masques, l'importance de notre santé mentale ou le rôle des médecins auprès des pouvoirs, l'intérêt d'un tel projet n'en était que plus fort.

C'est à partir du prisme historien que les différentes entrées de ce dictionnaire sont approchées : maladies, lieux, acteurs et actrices, concepts, événements, pratiques, parties du corps. En les resituant dans l'histoire et en insistant sur la dimension politique qui les traverse toutes, le dictionnaire contribue à dénaturer ces objets et à transformer l'image que le grand public peut en avoir. Les notices qui jusqu'à présent connaissent le plus grand succès sont, par exemple, celles qui permettent de penser la construction sociale des maladies féminines et du corps des femmes. À travers les discours des historiennes sur la *Folie puerpérale* (celle qui est liée à la grossesse et à la maternité), les discours sur l'*Hymen*, la pratique de la *Césarienne* ou la criminalisation de l'*Avortement*, le lectorat peut appréhender d'une manière accessible la dimension sociale et politique qui influence de longue date ces représentations du corps, ces pratiques et ces maladies.

Le projet, pensé en 2019 et réalisé en 2020, s'inscrit dans la politique scientifique du *Laboratoire Temps, Mondes et Sociétés* (TEMOS, UMR9016, CNRS/Université Angers/Université Bretagne Sud / Le Mans Université), au croisement de ses axes 1 - *Enfance, genre et traces de soi : individualités et subjectivités en mouvements* et 2 - *Ressources biologiques et construction des savoirs : circulations et usages*. Il s'est nourri du renouvellement considérable de l'historiographie francophone du champ de la santé perceptible également dans l'évolution de la revue *Histoire, médecine et santé*. Nombre des auteurs et autrices du dictionnaire ne se définissent pas comme des historiens et historiennes de la médecine *stricto sensu* mais comme des historiens et historiennes du social dont l'objet peut notamment être médical. Ils donnent désormais une définition plus large du champ de la santé qui ouvre les possibles pour écrire l'histoire d'acteurs et actrices moins connus (Associations de patients, entrepreneurs de santé), de pratiques considérées comme marginales (*Naturisme, Aérobic, Tatouage, Végétarisme*), de notions qui pourraient être jugées éternelles (*Urgence, Maladies rares*), de maladies oubliées et pourtant emblématiques de situations historiques pertinentes (*Polio, Drapétomanie*), de nouveaux concepts illustrant l'évolution permanente de nos conceptions de la maladie (*Neurodiversité, Schizophrénie*), de nouvelles disciplines et professions médicales qui illustrent le rôle central de la médecine dans nos sociétés (séries, films), de revendications citoyennes pour faire reconnaître officiellement certaines maladies et souffrances (sida, maladies du travail et de l'environnement), de notions liées aux handicaps également (surdité, cécité).



Cœur ailé et poignardé, tatoué sur la poitrine d'un récidiviste, photo du Laboratoire de police technique de Lyon, xx^e siècle. Source : B.I.U. Santé

Pour ce faire, le projet DicoPolHiS se déploie selon un format inédit qui combine les exigences du savoir académique et celles de la vulgarisation avec un objectif pédagogique original et une ouverture contributive. Les textes qui peuvent donc être produits par des chercheurs, chercheuses et universitaires de référence ou par des étudiants et étudiantes en formation (licence, master, doctorat) sont tous l'objet d'un même travail de réécriture pour présenter une cohérence à la lecture et pour atteindre le public visé. D'un maximum de 5 000 signes, correspondant à trois minutes de lecture environ, ces entrées sont toutes illustrées par une image d'archives, appuyées par deux références académiques et enrichies de nombreux liens hypertextes extérieurs ou renvoyant vers d'autres notices du dictionnaire. L'équipe du projet soutenu par le service communication de l'université du Mans comporte une quinzaine d'étudiants et étudiantes avancés des universités du Mans et d'Angers dont le travail se répartit en plusieurs pôles : édition, communication-réseaux, audiovisuel, partenariats, international. Portées par des réseaux sociaux dédiés, ces notices atteignent d'ores et déjà leur cible : un public de lecteurs et lectrices qui découvrent la dimension historique d'objets souvent considérés comme immuables et naturels. Citons ici quelques exemples.

En mettant l'accent sur des situations historiques méconnues du xix^e siècle, les notices *La Mecque*, *Livourne 1804* et *Barcelone 1821* offrent une longue perspective sur la gestion scientifique et politique des épidémies. À partir d'anciens épisodes de choléra et de fièvre jaune, les auteurs et autrices de ces textes illustrent la complexité des facteurs qui pèsent sur les prises de décision amenant les autorités à déclencher des processus de quarantaine ou à modifier leurs dispositifs. Les enjeux commerciaux se révèlent chaque fois prépondérants dans ces processus. Ceux-

ci s'entremêlent aux rumeurs et controverses scientifiques qui ne manquent pas de survenir à une époque où le principe de contagiosité peine encore à être consolidé dans les esprits des savants. Ces éclairages permettent de saisir les continuités et les discontinuités perceptibles entre les différentes pandémies que l'humanité a pu connaître. Les textes du dictionnaire éclairent également la relation entre médecins et patients et les racines de l'inégale répartition des soignants sur le territoire. L'entrée *Médecine à distance* lève le voile sur les pratiques de soin par correspondance au xviii^e siècle et celle sur les *Officiers de santé* rappelle l'époque durant laquelle des médecins de second rang prenaient en charge la santé des habitants des campagnes. L'entrée *Urgence* nous montre comment, après avoir été forgée par la culture du triage en temps de guerre, la lutte contre la mort accidentelle a été constituée peu à peu comme une priorité des administrations de la santé dans les sociétés industrielles.

L'historisation se révèle une démarche aussi pertinente dès lors qu'il s'agit d'identifier des maladies spécifiques à un genre. L'entrée *Folie puerpérale* montre comment un diagnostic psychiatrique s'impose progressivement dans la classification médicale alors que les symptômes qui le décrivent ne sont guère spécifiques. Mais, malgré cette faiblesse scientifique et les controverses qui l'accompagnent, le diagnostic survit. Aujourd'hui, certes, on utilise d'autres termes comme ceux de « dépressions périnatales » ou de *baby blues*, mais le consensus n'existe toujours pas sur ces diagnostics. Cet historique permet de nous questionner sur la part de construit et de naturalité dans les représentations sociales de la maternité. Il en va de même si l'on interroge la catégorie contemporaine de *Ménopause*.



Mazet dans une rue de Barcelone, lithographie de Jacques Arago, 1821, Wellcome collection





« Le marchand d'orviétan de campagne », gravure au pointillé en couleur, Louis Marin Bonnet, d'après Jacques Philippe Caresme, XVIII^e siècle

DicoPolHiS va bientôt entrer dans sa deuxième année et trois nouveaux développements vont être donnés au projet. En s'ouvrant aux collègues hispanophones et en créant une version anglaise du site au début de l'année 2021, le dictionnaire prend le tournant de l'international. Si, dans une première étape, des notices seront traduites en anglais, dans un second temps le site s'ouvrira aux contributions anglophones inédites. Un comité scientifique international en constitution prendra le relais, dans différents pays, de l'équipe francophone qui est aujourd'hui en charge de la recherche de sujets et d'auteurs et autrices. Le deuxième axe de réflexion porte sur la diversification des supports. Une équipe d'étudiants rompus à l'exercice radiophonique a été formée en décembre 2020 afin de proposer des adaptations originales du contenu du dictionnaire sous forme de podcasts audios. Une réflexion du même type se porte vers l'adaptation de notices sous format de brèves vidéos mêlant archives et discours historien. Enfin, le projet en est, aujourd'hui, au stade de la recherche de partenariats de publications avec des institutions

connectées au champ de la santé ou de la documentation historique. L'objectif est évidemment de pérenniser cet outil académique, pédagogique et contributif original afin d'envisager à terme de nouvelles formes de valorisation et de nouveaux chantiers de recherche collective.

Hervé Guillemain, professeur d'histoire contemporaine, Le Mans Université, TEMOS

► Retrouvez DicoPolHiS sur :   

contact&info

► Hervé Guillemain
TEMOS

Herve.Guillemain@univ-lemans.fr

► Pour en savoir plus

<http://dicopolhis.univ-lemans.fr/>



DicoPolHiS
Dictionnaire Politique d'Histoire de la Santé

Justice et intérêt - Aujourd'hui

Questionner les sciences économiques, leur pertinence, leurs limites, leur nature scientifique même, est une entreprise sans doute aussi ancienne qu'elles-mêmes (il suffit de relire les *Éthiques* et la *Politique* d'Aristote), mais dont l'urgence est rappelée à chaque crise — et c'est d'actualité. Dans la vie économique comme dans l'analyse que la science en fournit, les conflits entre intérêts divergents foisonnent en fonction d'exigences variées, la moindre n'étant pas la justice — quel que soit le plan envisagé : revenus, taxes et tâches à effectuer en société, des usines aux universités, des prisons aux hôpitaux, « en présentiel » comme « en distanciel » (selon des expressions devenues familières).

L'articulation de la justice et des intérêts place le savoir économique sur la sellette à la fois comme science au sein des sciences humaines sociales (SHS) et comme activité vitale de premier plan : dès lors qu'il s'agit de questions de survie (biologique et socio-économique), revendications et décisions sont à formuler dans le discours public comme dans les laboratoires de recherche. Ces derniers ont pour tâches une description des faits, l'énonciation des conditions de normes variées, l'émission de recommandations, voire de prescriptions ; quoique la décision relève d'ailleurs, déborde le cadre académique, les réflexions restent ancrées sur son socle. Or, la théorie permet effectivement d'envisager aussi la notion d'intérêt. Il ne faut voiler ni le jeu des intérêts dans la vie matérielle, ni les variations dans les conceptions de la justice. Des questionnements de philosophie économique portent en conséquence sur les notions de *justice* et d'*intérêt*. La réflexion s'élève contre les préjugés qui les entourent. Une critique philosophique digne de ce nom, qui vise un « tri », ou encore un juste départ entre bons et mauvais usages de la raison, peut et doit permettre un discours chargé de précautions, non d'imprécations. D'une part, la justice n'est pas la « douce rêverie » dont on croit parfois (à tort) que les philosophes se satisferaient (à trop bon compte). D'autre part, le rôle des intérêts matériels n'est ni négligeable ni méprisable : il fonde au contraire une bonne gestion des politiques publiques !

Le programme de travail du réseau international de recherche (IRN) « Justice & Intérêt » participe de cet effort. Il vise à le structurer, au sein de l'espace européen de la recherche, en réunissant neuf

centres européens de premier plan¹, en philosophie de l'économie (ou « philosophie économique », terme utilisé ici de préférence, et comme l'on parle de « philosophie morale, ou politique »)².

En reconnaissant l'importance de l'interdisciplinarité dans leurs pratiques, économistes théoriciens et philosophes coopèrent pour améliorer leurs théories, pour échanger des membres de leurs équipes pour des périodes courtes au long de quatre années d'IRN³. Ces chercheurs⁴ forment à la recherche par la recherche leurs collègues plus jeunes. Ils s'inscrivent dans une tradition déjà longue en épistémologie des SHS : on songera notamment à l'ouvrage de Jeanne Parrain-Vial, *La nature du fait dans les sciences humaines* (1966), aux travaux de Gilles-Gaston Granger, à certains textes politico-économiques (et avec cela méthodologiques) de Jules Vuillemin, tous deux professeurs à l'université de Provence (devenue en 2012 l'Université Aix-Marseille, qui abrite le laboratoire Aix-Marseille Sciences économiques). La philosophie économique fut marquée notamment par les personnalités d'Alain Leroux et de Pierre Livet, dont la collaboration fit surgir, il y a exactement vingt ans, la *Revue de philosophie économique* (bilingue, publiée à la Librairie philosophique Vrin). L'IRN est né sur ce sol aixois où la philosophie économique avait grandi depuis les années 1980, elle qui connaît un essor en accélération sensible (programmes universitaires, revues, etc.) depuis une dizaine d'années⁵.

La décision favorable de l'InSHS de créer l'IRN en 2018 a donné un rôle structurant aux unités françaises, autour de la méthodologie des économistes (une unité dépend de la section 37 - Économie et gestion, une autre de la section 35 - Sciences philosophiques et philologiques, sciences de l'art, la troisième est reconnue par une école doctorale dans chaque discipline). Pour des raisons liées à l'histoire des disciplines, la philosophie économique se pratique en France d'abord chez les économistes, et plutôt dans les départements de philosophie des sciences dans le monde académique influencé par le modèle anglo-saxon. La capacité à faire converser les membres actifs dans ces deux milieux, sur le fond d'arrière-plans conceptuels différents, est d'ores et déjà l'un des succès de l'IRN « Justice et intérêt ».

1. Pour dix tutelles, maximum autorisé pour un IRN. Les équipes européennes participantes, sans doute les plus réputées dans le domaine, sont les suivantes : le *Choice Group* du *Center for Philosophy of Natural and Social Sciences* (CPNSS), de la *London School of Economics* – LSE (Royaume-Uni) ; le *Center for Operations Research and Econometric* (CORE), Université Catholique de Louvain-la-Neuve (Belgique) ; le Centre Walras-Pareto (CWP), Université de Lausanne (Suisse) ; l'*Erasmus Institute for Philosophy & Economics* (EIPE), Université de Rotterdam (Pays-Bas) ; le laboratoire TINT, Université de Helsinki (Finlande).

Les équipes françaises impliquées sont : l'*Institut pour l'Histoire et la Philosophie des Sciences et des Techniques* (IHPST, UMR8590, CNRS / Université Paris 1 Panthéon-Sorbonne), Paris ; le laboratoire *Philosophie, Histoire et Analyse des Représentations Économiques* (PHARE, Université Paris 1 Panthéon-Sorbonne), Paris.

Enfin, le laboratoire *Aix-Marseille Sciences Économiques* (AMSE, UMR7316, CNRS / AMU) porte ce projet en la personne de l'auteur du présent article.

2. Sur le « choix des mots », voir l'ouvrage : Campagnolo G. et Gharbi J-S. 2017, *Philosophie économique. Une introduction*, Éditions Matériologiques, p. 57-80.

3. 2019 fut l'année de préfiguration et 2020 sa première année effective. Hélas, on s'en doute, malgré le « distanciel » et une liste d'informations dédiée au projet (« *irn_justice* »), ce programme de mobilité académique souffre évidemment de la pandémie. Une partie des réunions est maintenue à tout prix, comme les réunions annuelles.

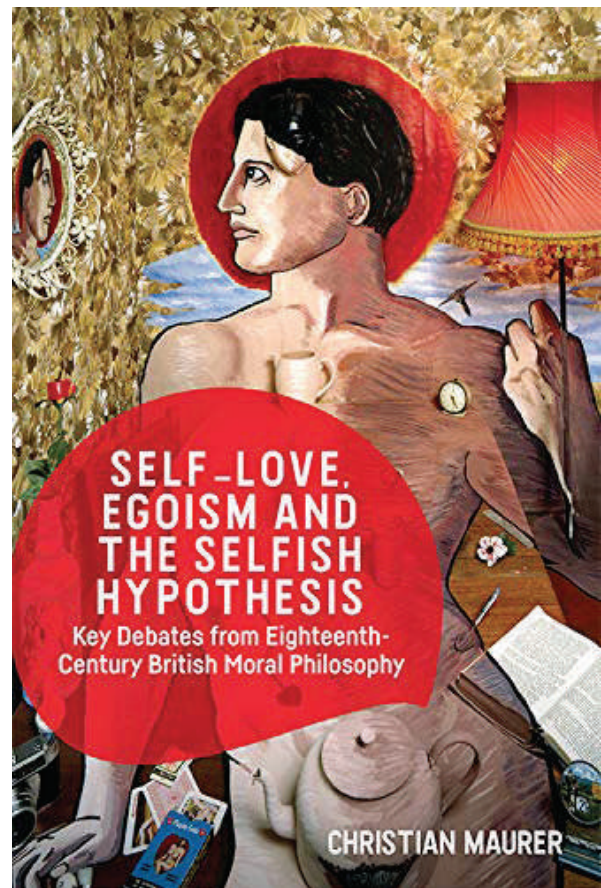
4. Plusieurs d'entre eux ont participé au collectif *Philosophie économique. Un état des lieux*, premier noyau structurant (volume de 618 pages coordonné par Gilles Campagnolo et Jean-Sébastien Gharbi aux Éditions Matériologiques).

5. Outre des conférences dont le rythme régulier s'établit dès 2010, des cursus universitaires ont émergé auxquels les unités de l'IRN contribuent chacune dans leur pays. On note l'essor des revues : *Economics and Philosophy*, *Journal of Economic Methodology*, *Revue de philosophie économique/Review of Economic Philosophy* (bilingue, créée à Aix-Marseille en 2000, co-dirigée par Gilles Campagnolo et Emmanuel Picavet). Plus récemment, citons l'*Erasmus Journal for Philosophy and Economics*, qui dépend du centre EIPE, tandis qu'*Economics and Philosophy* eut pour *Chief Editor* le *Team Coordinator* du *Choice Group*, Richard Bradley (CPNSS/LSE).

Parmi les travaux où la réflexion philosophique croise les enjeux majeurs de justice et d'intérêt, certains, comme les suivants, sont parfois passés dans le discours public : revenu de base⁶, dette publique⁷, moralité, équité et engagement économique⁸, éthique⁹ et questions de droits — comme la notion de « droit à la mobilité »¹⁰.

Donnons un exemple : l'économie du bien-être, qui s'appuie sur diverses théories de la justice, se diversifie en théories du choix social, théories de l'imposition et, inversement, de la (re)distribution/répartition (allocations équitables, soutien au développement humain, etc.) non seulement selon une analyse coût-bénéfice, mais aussi selon la mesure des inégalités et de la pauvreté, de la conception de mécanismes incitatifs (comme le *market design*) ou encore, de la fiscalité optimale. Si traiter toutes ces questions dépasse évidemment le périmètre d'un IRN, un apport de l'approche philosophique consiste précisément à saisir, sous la diversité des approches, les éventuels invariants pertinents à leur sujet. Le titre « Justice et distribution » (qui aurait pu s'appeler « Justice et répartition », l'anglicisme fut choisi à dessein) est le thème de la conférence annuelle 2020.

La philosophie économique qui s'étend ainsi aux « fondements de l'intervention publique » puise aux questions d'épistémologie concernant des notions de fond comme la nature de la *causalité*. Une véritable révolution épistémologique a eu lieu dans l'analyse causale (multivariée, multifonctionnelle, « contrôlée »), dont l'évolution des sciences économiques dépend (au moins en partie) : s'agit-il alors d'une normalisation de l'économie par rapport aux autres sciences, ou celle-ci garde-t-elle caractère d'exception, souvent revendiqué à l'intérieur et vilipendé à l'extérieur de celle-ci ? En philosophie économique, on y prête d'autant plus attention : ainsi, l'unité TINT (Helsinki) approfondit l'idéalisation de modèles applicables aux questions de justice dans les décisions économique¹¹. D'autres exemples sont : la rationalité (et/ou le caractère « raisonnable ») des décisions de politique économique, leur *design*¹², la probabilité/plausibilité des modèles mis en œuvre par les chercheurs en économie, induisant des effets dans l'application (ou pas) par les dirigeants, au sein d'institutions où les scientifiques apportent leur expertise¹³. En outre, dans la philosophie des sciences (où la philosophie économique se situe en milieu anglo-saxon), un débat actuel majeur porte sur l'*objectivité* de la connaissance, sa nature, ses limites, ce qu'elle implique pour la science. Là où Raymond Aron avait, en France, apporté de la clarté, les lignes sont redevenues confuses. Les tenants d'une forme d'épistémologie sociale voient la connaissance comme processus social construit et contraint de sorte qu'ils se rapprochent du relativisme (dans la foulée de philosophes comme Hilary Putnam), tandis que le mot de « diversité » implique encore d'autres considérations.



Couverture de l'ouvrage de Christian Maurer, *Self-Love, Egoism and the Selfish Hypothesis. Key Debates from Eighteenth Century British Moral Philosophy*

Sans que toutes les pistes ne puissent (ni ne doivent forcément) être suivies (cela dépend de leur pertinence aux yeux de chercheurs déterminant librement leur objet), il résulte de tout cela que la demande d'analyse philosophique des positions adoptées motive la réflexion (des critères d'allocations de ressources, selon la définition de l'économie de Lionel Robbins, aux incitations à des activités bénéficiant à la société ou, par exemple, l'éthique des populations¹⁴). La nature individuelle et/ou collective de la décision, dans un contexte économique mouvant, ainsi que la détection des atouts et des limites des résolutions prises trouvent leurs points décisifs dans les débats entre économistes et philosophes, issus d'une théorie de la décision désormais bien balisée¹⁵ et des fondements philosophiques de la conception de l'individualité, vue comme isolée ou en interaction avec autrui (responsabilité du « soi », « altruisme », conflits d'intérêt et enjeux de « confiance », délibération rationnelle et enjeux de court, moyen ou long terme des politiques, jugements portés et influencés par des normes sociales, etc.). Le dialogue entre économistes et philosophes est au cœur de ces démarches.

6. Voir, par exemple, l'article : Gamel C. 2018, « Fondements libéraux du revenu de base », *Revue de philosophie économique*, 19/2 : 61-92. Avant le début de l'IRN, l'auteur changea d'unité de recherche, de même que Jean-Sébastien Gharbi co-auteur avec Gilles Campagnolo des propositions d'IRN les années précédant son acceptation.

7. Voir la communication : de Saint-Phalle P. « La dette publique : histoire et normativité d'une institution politique moderne », Rencontre annuelle de l'IRN « Justice et intérêt », 14 décembre 2020.

8. Jaffro L., « Fairness, Impartiality, and Bargaining: Varieties of Moral Contractualism », colloque annuel des 19-20 septembre 2019.

9. Bradley R., « Social ethics under ambiguity », Rencontre annuelle de l'IRN « Justice et intérêt », 14 décembre 2020.

10. Dupont-Kieffer A., « The fantasy of mobility rights. Transportation and mobility in a fairer society », Rencontre annuelle de l'IRN « Justice et intérêt », 14 décembre 2020.

11. Voir ainsi : Aydinonat E., « Multiple model idealization in economics », Colloque de lancement de l'IRN, Septembre 2019, Marseille ; Mäki U., « Ideal and non-ideal theories of justice: confusions and clarifications », Rencontre annuelle de l'IRN « Justice et intérêt », 14 décembre 2020.

12. Voir ainsi : Van Basshuysen P., « Design economics: Past and future », Colloque de lancement de l'IRN, Septembre 2019, Marseille.

13. L'IRN est, pour l'heure, peu requis, mais l'expérience du porteur dans un projet antérieur (pour l'Union européenne) peut laisser envisager cela.

14. Pivato M., « Population ethics in an infinite universe », Rencontre annuelle de l'IRN « Justice et intérêt », 14 décembre 2020.

15. Voir l'ouvrage de Mikael Cozic à paraître aux Éditions Matériologique dans la collection « E-conomiques » (la collection est dirigée par Gilles Campagnolo).



Gilles Campagnolo, directeur de recherches au CNRS, porteur de l'IRN « Justice et intérêt » de l'InSHS, avec le Professeur Morimoto, organisateur de la conférence à l'université de Tokyo en hommage à Yuichi Shionoya

La mise en perspective historique est d'autant plus précieuse et l'IRN a permis de croiser les regards sur les conceptions de la justice qui s'expriment évidemment en rapport aux grandes théories de la justice, d'Aristote à Amartya Sen, en passant par le moment fondateur de l'économie politique dans le sensualisme des Lumières anglais comme français. L'économie politique passe par Smith¹⁶, et deux journées furent consacrées au thème de la « *Selfish Hypothesis* »¹⁷, ainsi que d'autres à John Rawls¹⁸ et Karl Popper¹⁹ notamment. À l'école des grands auteurs, mais avec les méthodes d'aujourd'hui, on refuse les caricatures réductrices — celles qui opposent, par exemple, la science mise en équations

et le souci du social, ou la quête de l'intérêt bien compris et l'aspiration à la justice. Ici, les enquêtes sur intérêt et justice vont de pair. Philosophie et économie manifestent à cet égard que chacune requiert l'autre. Il s'agit, en effet, de comprendre ce qu'est une vie socio-économique en même temps juste et consciente des intérêts qu'elle met en jeu. Afin, pour inverser un mot de Rimbaud, que la vraie vie soit présente et que nous soyons au monde. Les publications envisagées auront également ce but²⁰.

Un dernier mot : le travail de recherche en philosophie économique s'étend au-delà des frontières européennes du projet. Dans ce sous-champ disciplinaire et au-delà de l'Espace européen de la recherche (EER), que trouve-t-on en effet ? En Amérique du Nord, on dénombre malheureusement fort peu d'unités travaillant sur ce champ²¹. En Asie orientale, existe — au Japon — une tradition relativement ancienne (puisque née dans l'Entre-deux-guerres) qui reste brillamment illustrée : citons notamment les travaux de Yuichi Shionoya (1932-2015). Il est souhaitable que le projet s'étende au Japon. Or, des études communes le permettent²², visant un rapprochement à venir.

contact&info

► Gilles Campagnolo
Amse

gilles.campagnolo@univ-amu.fr

16. Voir ainsi la communication : Bee M., « Adam Smith's Third Concept of Justice », Colloque de lancement de l'IRN, Septembre 2019, Marseille.
17. Notamment les 25 janvier et 4 mai 2020 à Paris, autour du livre : Maurer C. 2019, *Self-Love, Egoism and the Selfish Hypothesis. Key Debates from Eighteenth Century British Moral Philosophy*, Edinburgh University Press.
18. Claude d'Aspremont (UCLouvain) discuta les « préférences fondamentales » et Feriel Kandil (AMSE) compara Ricoeur et Rawls quant à l'« aporie du juste » (publié en anglais) lors de la réunion de lancement de septembre 2019.
19. Conférence donnée le 2 mars 2020 à la *London School of Economics* par Gilles Campagnolo, touchant son édition des *Écrits de jeunesse* de Popper à Vienne (Éditions rue d'Ulm, 2019).
20. Notamment un numéro thématique de la *Revue de philosophie économique* (bilingue) et, possiblement, un recueil en langue anglaise.
21. Il existe le *Center for the History Of Political Economy* (CHOPE, université Duke, Caroline du Nord), mais il adhère étroitement à son intitulé, et le Centre de philosophie des sciences à Pittsburgh fait bien peu d'économie.
22. Une conférence, organisée le 21 décembre 2019 à l'université de Tokyo, est venue couronner le numéro spécial « Philosophie économique au Japon » de la *Revue de philosophie économique* (20/1, juin 2019) où des inédits de Yuichi Shionoya en toute autre langue occidentale que le français ont pu paraître (droits obtenus par Gilles Campagnolo, offerts par les héritiers).

Une biographie collective au xx^e siècle, entre histoire des migrations et histoire de la Shoah

Directrice de recherche CNRS au sein de l'Institut d'histoire moderne et contemporaine (IHMC, UMR8066, CNRS / ENS Paris / Université Paris 1 Panthéon-Sorbonne), l'historienne Claire Zalc a obtenu, en 2018, un financement ERC Consolidator Grant pour le projet « LUBARTWORLD - Migrations et Shoah : les trajectoires transnationales des Juifs de Lubartów à travers le monde, années 1920-1950 ». L'objet de ce projet est d'interroger les liens entre dynamiques migratoires et persécution. L'enjeu consiste à étudier à la fois ceux qui restent et ceux qui émigrent dans différents espaces tout en interrogeant les connexions entre ces deux groupes.



Groupe de jeunes juifs, Lubartow, années 1930 © Yad Vashem

Pourquoi avez-vous postulé à l'ERC ?

Il m'a fallu presque dix ans pour élaborer et mûrir ce programme de recherche, qui me tient particulièrement à cœur. L'idée en est simple, à première vue : tenter de reconstituer une par une les trajectoires biographiques de l'ensemble des habitants juifs d'une petite bourgade de Pologne, Lubartów, des années 1920 aux années 1950, qu'ils aient émigré ou qu'ils soient restés sur place, qu'ils aient été exterminés ou qu'ils aient survécu à la Shoah. J'ai commencé à y travailler seule, mais me suis rapidement aperçue que j'avais besoin, pour le mener à bien, de postuler à un financement ERC. Pour trois raisons principalement. Tout d'abord, je ne possède pas l'ensemble des compétences linguistiques nécessaires, puisque l'enquête compte des documents aussi bien

en polonais, en russe, en yiddish, en hébreu, en espagnol, en allemand, en portugais, en français qu'en chinois et en japonais. Ensuite, la dispersion des sources sur plusieurs continents nécessite la réalisation de fréquentes missions d'archives, longues et dans un nombre conséquent de pays, que seul le type de financement proposé par l'ERC pouvait rendre possible. Enfin, ce projet appelle une véritable réalisation collective, car il réunit une importante partie de collecte archivistique, de travail sur les images et les sources photographiques, mais également de quantification poussée du matériel recueilli, et de transmission au grand public à travers la réalisation d'un site Internet notamment, de guides de recherche et également, peut-être, d'exposition et de documentaire. La dimension européenne de mon projet étant



From Lubartow to the world : le projet Lubartworld © Emeline Néant

indéniable, j'ai forcé ma nature, ma culture et ma formation à me conformer à ce défi, sans trop y croire, la première année, pour être honnête. J'ai d'ailleurs d'abord essuyé un échec, après un oral difficile la première année qui m'a sans doute aidé à repenser certains aspects du projet, sa formulation et sa présentation aux membres du comité, pour réussir l'année suivante.

En quoi la méthode originale de votre projet (le suivi longitudinal des trajectoires des juifs d'une petite ville de Pologne sur trois ou quatre décennies) est-elle de nature à éclairer d'un jour nouveau à la fois l'histoire des migrations et celle de la Shoah ?

L'histoire des migrations et l'histoire de la Shoah sont deux historiographies en plein renouvellement qui pourtant s'ignorent largement. L'objectif principal du projet consiste à comprendre les dynamiques d'une structure sociale soumise à une perturbation majeure, en étudiant les conditions sociales et les conséquences de la destruction d'un collectif. Qui fuit ? Quand ? Où aller ? Avec qui ? Qui survit et qui ne survit pas ? Il ne s'agit pas seulement de compter combien d'individus sont partis ou sont restés, ont été déportés ou non, mais bien de comprendre les liens entre les trajectoires de persécution et les trajectoires de migration à la lumière des environnements familial, économique et de voisinage. Ce faisant, le projet permet d'évaluer comment la chronologie de la mise en place des persécutions pèse, ou pas, sur les dynamiques et comportements migratoires, mais aussi les espaces de destinations. Nous avons déjà repéré une vingtaine de pays d'émigration. En procédant de manière exhaustive au sein d'un groupe, le projet se donne les moyens de comparer les itinéraires. Quel est le rôle joué par le statut socio-économique, le sexe, la taille de la famille, le degré de religiosité ou encore l'appartenance politique dans les manières diverses de se positionner face aux persécutions ? Quelles explications peut-on proposer pour rendre compte des différences de destinations migratoires ?

Concernant l'histoire de la Shoah, ce projet contribue à montrer l'importance de la désenclaver pour la considérer à la fois dans ses dimensions transnationales et dans une certaine épaisseur chronologique, que je qualifie de biographique. Le projet s'étend du milieu des années 1920 au milieu des années 1950, des conséquences de la Première Guerre mondiale, avec la disparition des empires multinationaux, aux reconfigurations territoriales, politiques et identitaires consécutives à la Seconde Guerre mondiale et la Shoah. Ceci conduit à relire les rythmes de cette histoire, à l'échelle non plus des décideurs mais des victimes, dispersées sur un ensemble de territoires, bref, dans son épaisseur biographique. Il s'agit d'explorer cette séquence chronologique du point de vue des personnes qui vivent tant les épisodes migratoires que les étapes des persécutions antisémites. C'est pourquoi, si le projet commence à Lubartów, il mène au monde dans sa globalité puisque les Lubartoviens le sillonnent et que leurs trajectoires incarnent à leur manière les bouleversements que connaissent l'Europe et le monde au ^{xx}e siècle, comme la création et la destruction des frontières et d'États-nations, les massacres et les politiques de reconstruction ou de réinstallation après la Seconde Guerre mondiale.

En outre, en adoptant l'échelle d'un groupe de plus de 3 000 personnes, ce projet engage à l'exploration de techniques quantitatives originales. La mise au point de bases de données relationnelles, puisque plusieurs dizaines de corpus de sources sur les mêmes individus sont mobilisés, constitue un défi méthodologique. Enfin, la dimension d'histoire orale et de recueil de témoignages filmés du projet — puisque j'ai été très rapidement mise en contact avec les très rares survivants de Lubartów et surtout leurs enfants — met en lumière des éléments tout à fait contemporains des manières d'écrire cette histoire, alors que les témoins ne seront bientôt plus là pour en parler.

Quels conseils donneriez-vous aux chercheurs qui souhaitent se lancer dans la préparation d'un ERC Consolidator Grant ?

Mon premier conseil serait de les inciter à ne pas s'autocensurer, en rappelant qu'il y a, là, des opportunités de financement qui sont possibles à obtenir. Le deuxième conseil, corollaire du premier, les inviterait à considérer la phase de préparation et de soumission comme un moment de recherche en soi. Ayant vécu à la fois la déception de ne pas voir son projet retenu, puis le succès, je crois que la meilleure manière d'aborder cette épreuve consiste à en faire une rampe de lancement pour un projet auquel on tient vraiment, en l'améliorant. Penser aux questions de faisabilité, aux formats et aux chemins que l'enquête pourra prendre, aux compétences dont on a besoin pour le mener à bien, sont autant d'avancées dans la maturation du projet. Il me semble néanmoins qu'il faut rappeler que l'obtention d'un financement ERC est un moyen et non une fin en soi. C'est pourquoi j'ai l'impression qu'il vaut mieux présenter un projet qui nous tient à cœur, et sur lequel on planche et rêve depuis longtemps, parce que la phase de soumission permet de l'améliorer, de le faire avancer et devient, du coup, presque agréable. Je dis bien « presque » car les montages de budgets, les batailles avec les contraintes notamment des sites internet de l'ERC, et autres formalités bureaucratiques ne doivent pas non plus être ignorées. Sur ce point, j'ai eu la très grande chance d'être aidée par les équipes formidables d'aide au montage de projets de ma Délégation CNRS, mais aussi de mes deux institutions de rattachement, l'ENS

et l'EHESS. Cela m'a été extrêmement utile de pouvoir compter sur leurs compétences. J'ai aussi sollicité d'anciens lauréats et lauréates, des collègues, des amis et pu bénéficier de leurs conseils. Il ne faut pas hésiter à demander de l'aide !

Enfin, j'aimerais rappeler que l'obtention d'un ERC n'est pas uniquement une panacée. Elle s'accompagne aussi de difficultés de tous ordres, qu'il ne faut pas passer sous silence, et de contraintes. C'est aussi épuisant, notamment lorsqu'il s'agit de mener un projet transnational au milieu d'une crise sanitaire, avec archives, bibliothèques et frontières fermées...

Penser l'impensable, anticiper l'inattendu, être autant scientifique, gestionnaire, chef d'entreprise, déménageur que chercheur. « Agir en primitif, prévoir en stratège » (René Char, *Feuillets d'Hypnos*, n° 72, Paris, Gallimard, 1962).

contact&info

► Claire Zalc
IHMC

Claire.Zalc@ens.fr

► Pour en savoir plus
<https://lubartworld.cnrs.fr/>



Commémoration des amis de Lubartow, cimetière de Bagneux, France années 1960 © Mémorial de la Shoah

Linguistique Computationnelle

La linguistique computationnelle, comme plus généralement la linguistique, cherche à comprendre le fonctionnement du langage humain. Mais sa particularité est d'aborder cette question dans une perspective de calcul informatisé, c'est-à-dire en développant des algorithmes qui peuvent être exécutés par un programme informatique pour simuler ou modéliser notre capacité humaine à percevoir, comprendre, produire le langage. Autrement dit, la linguistique computationnelle s'applique à découvrir les systèmes de traitement qui permettent de percevoir, comprendre, générer, le langage naturel — qu'il s'agisse des sons, des mots, des combinaisons de mots, des énoncés (en nombre potentiellement infinis), individuels ou structurés en discours, textes ou dialogues, et cela, par principe, dans n'importe quelle langue donnée et quelle que soit la modalité (orale, écrite, gestuelle), ou même dans plusieurs langues puisque nous pouvons comprendre ou parler plusieurs langues à la fois. Elle cherche à atteindre ces objectifs à travers des activités qui reposent sur le développement d'outils informatiques, et l'utilisation de méthodes symboliques (approches logiques, systèmes à base de règles), de méthodes statistiques basées sur des techniques d'apprentissage automatique permettant d'extraire des connaissances à partir de données ou de méthodes hybrides symboliques/statistiques.

Il s'agit donc là d'un champ de recherche par essence interdisciplinaire combinant les ressources de toutes les branches de la linguistique (phonétique, phonologie, morphologie, syntaxe, sémantique, pragmatique, sociolinguistique...), de l'informatique, des mathématiques, de la logique, de la modélisation statistique, et s'inscrivant à la croisée des différents champs scientifiques qui s'intéressent à l'étude de la faculté de langage et aux autres capacités cognitives (raisonnement, perception, décision, planification...), ainsi qu'à leur modélisation : sciences du langage, sciences de l'information, intelligence artificielle, philosophie du langage et de l'esprit, anthropologie, sciences cognitives (psychologie, psycho- et neuro-linguistique, neurosciences, notamment).

La linguistique computationnelle naît dans le sillage de mouvements scientifiques qui se développent dans les années 1950 : les recherches en traduction automatique (à l'essor desquelles le CNRS contribuera en France en créant, en décembre 1959, le Centre d'Études sur la Traduction Automatique (CETA) au sein du Laboratoire de calcul numérique de l'Institut Blaise-Pascal du CNRS¹), et le tournant cognitiviste² qui cherche à caractériser l'esprit humain et la connaissance humaine en faisant appel aux concepts et aux techniques qui se développent avec l'émergence de l'intelligence artificielle, de l'informatique et des neurosciences.

La linguistique computationnelle a des motivations et des objectifs qui relèvent autant de la recherche fondamentale que de la recherche appliquée. Recherche fondamentale lorsqu'il s'agit, par exemple, de développer une explication/ un modèle computationnel d'un phénomène linguistique ou psycholinguistique — comme le processus d'acquisition du langage chez les enfants —, ou encore de tester, de mettre à l'épreuve des modèles linguistiques ou des hypothèses psycholinguistiques sur le traitement du langage. Recherche appliquée lorsqu'il s'agit de créer des outils permettant d'accomplir des tâches qui, à l'ère du numérique, prennent de plus en plus d'ampleur, comme, par exemple, la traduction automatique, la recherche et l'extraction automatique d'information à partir de textes, la correction orthographique et grammaticale, les systèmes de question/réponse, le dialogue homme-machine ou la reconnaissance et la synthèse vocales.

On ne saurait faire honneur dans les quelques pages de ce zoom à la diversité et à la richesse des travaux conduits en linguistique computationnelle à l'InSHS à travers ses chercheurs, chercheuses et ses unités, en étroite collaboration avec l'Institut des sciences de l'information et de leurs interactions - INS2I et l'Institut des sciences biologiques - INSB (principalement). On espère néanmoins en illustrer quelques aspects, notamment en interface avec la linguistique de terrain et les défis que soulèvent les langues non-documentées ou sous-dotées, ainsi que la question plus générale de la préservation de la diversité linguistique, la linguistique historique et typologique, la linguistique formelle, la linguistique expérimentale ou les neurosciences.

Hamida Demirdache, DAS InSHS

1. Léon J. 2002, Le CNRS et les débuts de la traduction automatique en France, *La revue pour l'histoire du CNRS*, Comité pour l'histoire du CNRS 6 : 6-24.

2. L'un des événements qui marque la naissance des sciences cognitives est la tenue d'un symposium sur la théorie de l'information en septembre 1956 au *Massachusetts Institute of Technology*, où sont présentés des travaux en psychologie, linguistique et informatique. Comme le rappelle George Miller : "I went away from the Symposium with a strong conviction, more intuitive than rational, that human experimental psychology, theoretical linguistics and computer simulation of rational cognitive processes were all pieces of a larger whole, and that the future would see progressive elaboration and coordination of their shared concerns" (Miller G.A. 2003, *The Cognitive Revolution: a Historical Perspective*, *Trends in Cognitive Sciences* 7(3) : 141-144.

Linguistique Informatique: historique, applications et modélisation

Claire Gardent est directrice de recherche au sein du *Laboratoire lorrain de recherche en informatique et ses applications* (LORIA, UMR7503, CNRS / Université de Lorraine / Inria). Elle s'intéresse aux modèles hybrides statistiques/symboliques pour la modélisation informatique des langues naturelles. Elle travaille sur l'acquisition automatique de grammaires et de ressources lexicales, sur l'analyse syntaxique et sémantique, sur la génération automatique de textes et sur les technologies innovantes pour l'apprentissage des langues.

Quelle peut bien être la relation entre langue et informatique ? Cette alliance, qui peut paraître surprenante, a aujourd'hui de nombreuses applications. C'est aussi un domaine de recherche qui contribue au renouvellement des méthodes utilisées en linguistique formelle et en linguistique de terrain.

Des débuts faussement prometteurs

C'est la traduction automatique qui, dans les années 1950, a initié ce domaine de recherche. Selon le mémorandum Weaver de 1949, la traduction automatique pouvait être vue comme une opération de décodage et s'inspirer des méthodes de décryptage utilisées par les ordinateurs pendant la Seconde Guerre mondiale. Il suffisait de stocker un dictionnaire bilingue sur un ordinateur puis de traduire mot à mot.

Rétropédalage

La complexité du langage humain a rapidement montré les limites d'une telle idée et l'approche a échoué entraînant le gel des financements pour la recherche en linguistique informatique. Peu à peu cependant, de nouvelles recherches ont émergé qui visaient, non plus à produire une application (un traducteur automatique), mais à modéliser les différents aspects de la langue. Quelles sont les formes possibles d'un mot (morphologie) ? Quelles sont les règles qui régissent leur combinaison dans une phrase (syntaxe) ? Quelles sont celles qui permettent de passer du texte au sens (sémantique) et de prendre en compte le contexte (pragmatique) ? Quels sont les sons d'une langue et comment se combinent-ils (phonétique et phonologie) ? Instruits par l'échec de la traduction automatique, dans les années 1960-1980, les linguistes informaticiens ont cherché à développer des modèles informatiques qui rendent compte de la complexité de la langue et des multiples réseaux de contraintes qui la gouvernent. Ils se sont en particulier concentrés sur le développement de grammaires électroniques et d'algorithmes d'analyse syntaxique permettant de produire la structure syntaxique d'une phrase. De larges ressources électroniques ont également été créées, telles que WordNet qui décrit les relations de synonymie entre les mots et peut être utilisé pour déterminer si deux phrases ont un sens similaire.

Vers la fin des années 1980, ces travaux se sont heurtés à une caractéristique première de la langue : l'ambiguïté. Les sons, les mots, les structures de la langue sont fréquemment ambigus. Un son peut correspondre à plusieurs graphies (« son », « sons », « sont »). Un mot peut avoir plusieurs catégories (« livre » peut être un verbe ou un nom) et plusieurs sens (le nom « livre » peut désigner l'objet à lire ou la monnaie). Une phrase peut avoir plusieurs interprétations. Ces ambiguïtés entraînent une explosion combinatoire (un ensemble très large de solutions possibles) que les algorithmes développés ne pouvaient ni traiter, ni résoudre. Parallèlement, les recherches en traitement de la parole et les recherches d'information ont fait émerger l'utilisation de méthodes statistiques permettant de calculer la

probabilité d'une solution et, ainsi, de choisir — parmi l'ensemble des solutions produites par un modèle — la solution la plus probable. En linguistique informatique, ces méthodes ont permis le développement d'approches dites statistiques qui permettent, à partir d'un corpus annoté, c'est-à-dire de textes ou de phrase enrichis par des experts avec des informations linguistiques, de calculer puis d'utiliser des probabilités pour lever les ambiguïtés de la langue. L'idée clé capturée par ces nouvelles approches est que la langue est régie par un ensemble de contraintes dont l'interaction peut être modélisée par des probabilités. Ainsi, les analyseurs syntaxiques statistiques pour l'anglais entraînés sur le corpus arboré de l'université de Pennsylvanie (un corpus de 40 000 phrases extraites du *Wall Street Journal* et annotées chacune avec leur arbre syntaxique par des linguistes) permettent de créer automatiquement l'arbre syntaxique d'une phrase très rapidement et avec une grande précision.

Plus récemment, les approches neuronales se sont largement imposées parce qu'elles permettent de créer des représentations linguistiques de bonne qualité et, plus généralement, parce que les résultats obtenus par ces méthodes sur les problèmes et les applications de la linguistique computationnelle dépassent souvent, de loin, ceux des autres méthodes statistiques.

Aujourd'hui, un domaine en pleine expansion

Le domaine qui, dans les années 1980, rassemblait quelques centaines de délégués dans ses grandes conférences, a vu une véritable explosion de sa communauté scientifique avec, aujourd'hui, des conférences rassemblant plusieurs milliers de délégués venus non seulement des universités du monde entier mais aussi, et de façon massive, des centres de recherches du privé, notamment des GAFAM (Google, Amazon, Facebook, Apple, Microsoft).

Les thèmes de recherche de ces conférences et, plus généralement, de la linguistique informatique sont multiples et variés.

Certains concernent la formalisation informatique du langage. Un objectif clé du domaine est de développer des méthodes automatiques qui permettent de dériver d'une phrase ou d'un texte, une caractérisation de sa structure linguistique (phonétique, phonologique, morphologique, syntaxique ou sémantique). On cherche également à créer (de façon automatique) des représentations du sens des mots qui rendent compte des relations de synonymie ou d'antonymie entre les mots d'une même langue ou de langues différentes ; on étudie les différences et ressemblances entre langues d'une même famille ou de familles différentes ; ou encore, on simule l'acquisition du langage, c'est-à-dire la façon dont l'humain apprend à parler et à comprendre la langue.

D'autres thèmes sont plus applicatifs. Ainsi, une part importante de la recherche en linguistique informatique consiste à concevoir des modèles de dialogue homme-machine, de traduction

automatique ou de question-réponse. Même si ces modèles sont aujourd'hui présents dans nos vies sous la forme d'applications que nous utilisons au quotidien — tels les agents conversationnels, les traducteurs automatiques ou les assistants personnels —, ces produits sont souvent restreints dans leur utilisation (utilisables seulement pour un domaine ou pour une langue donnée). Un objectif important des recherches actuelles est d'améliorer leur portabilité, c'est-à-dire de faciliter leur portage à un domaine ou à une langue nouvelle. Les recherches visent également à développer des modèles qui puissent prendre en compte différentes sources d'information (parole, texte, image, signal) ou encore à concevoir des méthodes qui permettent de simplifier ou de résumer un ou plusieurs documents.

Enfin, des thèmes ont récemment émergé qui découlent de considérations sociales et reflètent la place grandissante des applications et techniques développées par le domaine dans notre société : comment développer des modèles « verts » qui aient une empreinte carbone réduite ; comment produire des modèles ou des systèmes éthiques, dépourvus de biais et de comportement sexistes, racistes ou offensifs ; comment exploiter les techniques développées pour faciliter l'accès à la culture et au savoir (humanités numériques), pour générer ou interpréter la langue des signes ou pour permettre le traitement informatique des langues dites « peu dotées », c'est-à-dire des langues pour lesquelles peu d'applications existent.

Linguistique informatique, linguistique formelle et linguistique de terrain

La linguistique informatique a des liens étroits avec la linguistique formelle et la linguistique de terrain. Ainsi, les analyseurs syntaxiques et les grammaires électroniques permettent de confronter modèle formel et données empiriques au sens où ils permettent de valider ou d'invalider un modèle sur un nombre illimité d'exemples (tirés du web ou d'un corpus de textes), plutôt que sur quelques exemples construits par le linguiste pour illustrer les prédictions de son modèle. Le développement et la mise à disposition de larges corpus textuels, de concordanciers¹ permettant de fouiller ces corpus de façon ciblée ou encore de larges corpus arborés associant chaque phrase d'une langue avec

une description de sa structure syntaxique, sont également des ressources précieuses pour analyser les structures de la langue ou pour extraire des généralisations linguistiques. Les ressources multilingues annotées par les linguistes informaticiens peuvent notamment être utilisées, en combinaison avec des méthodes d'apprentissage, pour assister les typologues dans le travail qui consiste à classifier les propriétés des systèmes linguistiques ou à dresser une typologie des langues du monde.

Les outils et les techniques développés par la linguistique informatique peuvent également contribuer à accélérer la collecte des données par les linguistes de terrain et à produire des données de meilleure qualité pour l'analyse linguistique. Traditionnellement, le linguiste de terrain documente les langues du monde par une interaction avec les locuteurs natifs en enregistrant leur parole et en leur posant des questions. Il lui faut ensuite transcrire ces données (passer de l'oral à l'écrit), les traduire et les analyser. Depuis peu, une collaboration croissante entre linguistique de terrain et linguistique informatique vise à accélérer et à améliorer ce processus de collecte, de traitement et d'analyse. On cherche, en particulier, à adapter les techniques de reconnaissance vocale, d'alignement texte/audio et de traduction automatique développée par la linguistique informatique aux langues orales non ou peu documentées et étudiées par les linguistes de terrain. Cet axe de recherche a toute son importance à une époque où de nombreuses langues sont en voie de disparition. On estime actuellement que sur environ 6 000 langues parlées dans le monde, près de 3 000 seraient en voie d'extinction soit parce que leurs locuteurs disparaissent soit parce qu'ils adoptent la langue dominante. Dans ce contexte, la linguistique informatique a un rôle important à jouer : elle peut contribuer à documenter un patrimoine culturel en voie d'extinction.

contact&info

► Claire Gardent

Loria

claire.gardent@loria.fr

1. Un concordancier est un logiciel qui permet d'extraire d'un corpus un mot et son contexte permettant ainsi d'en étudier l'usage.

Outiller les langues régionales : expériences coopératives sur l'occitan et l'alsacien avec l'aide du français, de l'allemand, du serbe, du catalan...

Professeure de linguistique, Myriam Bras mène des recherches dans le domaine de la sémantique. Sa démarche allie description, formalisation et construction de ressources pour la linguistique occitane. Aleksandra Miletic est spécialiste de traitement automatique des langues et de la constitution de ressources pour les langues peu dotées. Chercheur CNRS en linguistique, Jean Sibille est spécialiste des langues régionales et plus particulièrement de la langue occitane. Dejan Stosic est maître de conférences HDR en sciences du langage. Ses recherches se situent à l'interface de la sémantique et de la syntaxe. Il est spécialiste de la sémantique de l'espace, de la polysémie et de la comparaison des langues. Tous les quatre sont affiliés au laboratoire *Cognition, Langues, Langage, Ergonomie* (CLLE, UMR5263, CNRS / Université Toulouse Jean Jaurès / Université Bordeaux Montaigne). Membre du laboratoire *Formes et Représentations en Linguistique, Littérature et dans les arts de l'Image et de la Scène* (FoReLLIS, EA 3816, Université de Poitiers), Marianne Vergez-Couret inscrit ses travaux dans le champ de la sémantique du discours et participe au développement de ressources et outils pour l'analyse automatique de l'occitan et du poitevin-saintongeais. Delphine Bernhard est maîtresse de conférences en informatique au sein du laboratoire *Linguistique, Langues, Parole* (LiLPa, UR 1339, Université de Strasbourg) et spécialiste du traitement automatique des langues et en particulier de la constitution de ressources et outils pour les langues peu dotées.

(In)égalité des langues

Les langues du monde ne sont pas sur un pied d'égalité, qu'il s'agisse de leur existence normale de langues parlées par des groupes d'humains, ou des efforts de ces mêmes humains pour les préserver comme autant de joyaux de la culture de l'humanité. Ce triste adage est hélas vérifié aussi en linguistique, où, contrairement à ce que l'on pourrait imaginer, on peut dire sans exagérer qu'une majorité de linguistes travaillent sur une minorité de langues. Et ce, malgré les efforts des typologues, des linguistes de terrain et de linguistes conscients de la richesse de la diversité des langues du monde, qui font de leur mieux pour préserver, étudier et faire connaître ce patrimoine de l'humanité¹. La tendance est encore plus accrue dans le domaine du « traitement automatique des langues » et de la « linguistique informatique ou computationnelle » ; ces champs de recherche ont ouvert la voie aux « technologies du langage » dont il est aisé de comprendre l'extrême sensibilité aux pressions économiques et politiques qui vont de pair avec une vision utilitariste des langues, très répandue dans nos sociétés. Il n'est donc pas étonnant que la présence ou l'absence des langues dans l'univers numérique d'internet et des réseaux sociaux reflète ces inégalités et que le numérique creuse de façon dramatique les écarts entre les langues, plaçant l'anglais seul en tête sur la toile mondiale, très loin devant toutes les autres langues du monde².

Heureusement, un nombre croissant de chercheuses et chercheurs envisagent les technologies du langage comme des outils pour explorer la diversité linguistique et faire vivre les langues, s'inscrivant ainsi à contre-courant de la disparition programmée d'un grand nombre d'entre elles³. Nous en faisons partie et reprenons à notre compte l'adage plus optimiste de Félix Marcel Castan, emblématique du *Forum des Langues du Monde* de Toulouse, « les langues et les cultures du monde sont égales entre elles comme les citoyens d'une même république », pour l'appliquer au domaine de la linguistique computationnelle. Nous

avons choisi de travailler sur des langues vivantes — parlées, chantées, écrites, enseignées, véhicules de créations littéraires et artistiques — reconnues comme appartenant au patrimoine de la France⁴, l'occitan et l'alsacien. Nous avons choisi d'unir nos efforts pour faire sortir ces langues de l'état de « langues peu dotées » en ressources numériques et outils de traitement automatique, en nous appuyant sur les avancées de la recherche en linguistique computationnelle sur les langues « mieux dotées » comme l'anglais, le français ou l'allemand, et sur des expériences récentes d'outillage d'autres langues « moins dotées » comme le serbe et le catalan. Nous proposons dans cet article de partager les enjeux et les défis liés à l'outillage de ces langues ainsi que les réalisations et les enseignements que nous tirons de ces expériences.

Outiller les langues : enjeux et défis

Outiller les langues, c'est créer des ressources lexicales et textuelles ainsi que des outils d'analyse automatique pour aider les linguistes qui cherchent à les décrire, leur permettant ainsi de décupler l'accès à leurs objets scientifiques. Les ressources en question vont des lexiques informatiques aux corpus de textes, d'abord des corpus bruts puis des corpus annotés, c'est-à-dire enrichis avec des informations linguistiques qui seront autant de points d'entrée dans les données (catégories grammaticales des mots, informations sur le genre, le nombre, la personne, le temps, relations syntaxiques entre les mots, ou encore divers types d'informations sémantiques). Les lexiques et les corpus de textes — écrits ou oraux — ainsi enrichis, organisés en bases textuelles monolingues ou multilingues, sont des ressources précieuses pour les chercheurs en linguistique descriptive, en linguistique théorique, en linguistique contrastive, et en typologie. Si l'on considère que chaque langue présente un intérêt sur le plan scientifique, on comprend aisément l'enjeu de cette entreprise pour faire progresser la connaissance des langues de l'humanité.

1. Voir par exemple le [programme SOROSORO](#) de l'association WOLACO (Conservatoire des langues du monde).

2. À cela s'ajoutent les critères de publication et d'évaluation par les pairs en TAL qui mettent l'accent sur la comparaison avec les méthodes existantes, favorisant ainsi la réutilisation des mêmes jeux de données (très souvent en anglais) pour l'évaluation des méthodes, et une recherche effrénée de performances parfois au détriment de la diversité et de la qualité des données.

3. Voir, par exemple, l'émergence de conférences ou d'ateliers dédiés comme TALaRE (Traitement Automatique des Langues Régionales de France et d'Europe) associé à TALN (Traitement Automatique des Langues Naturelles) en France ou VarDial associé à COLING (*International Conference on Computational Linguistics*) au plan international, et le focus sur les langues en danger de LREC (*Language Resources and Evaluation Conference*) 2020 ou LT4All (*Language Technologies for All*).

4. Mais ne bénéficiant d'aucun statut officiel en France — où la Charte Européenne des Langues Régionales et Minoritaires n'a pas été ratifiée — contrairement à l'Italie et l'Espagne, pour ce qui concerne l'occitan.

Rudyard Kipling: Knjiga o džungli — 1894

a kad je granulo jutro i magla bila sasvim bela i hladna, primetio je kako je seljakov dečak uzao jedan trščani lonac iznutra oblepljen biatom, u njega spustio komadiće zažarenog ugljevlja, pokrio ga platnom i izišao napolje da vidi krave u staji. — Je li to sve? — reče Mogli. — Ako to može da radi čovekovo mlađunče, stvar nije opasna.

Rudyard Kipling: The Jungle Book — 1894

And when the morning came and the mists were all white and cold, he saw the man's child pick up a wicker pot plastered inside with earth, fill it with lumps of red-hot charcoal, put it under his blanket, and go out to tend the cows in the byre. "Is that all?" said Mogli. "If a cub can do it, there is nothing to fear."

Rudyard Kipling: Mowgli, la Granhòta — 1894

Quand puèi se sarrèt lo matin, a l'ora que s'embianquesis la nèbla freja, vegèt lo pichon de l'òme prèner un descon de vims perbocat de rìòla en dins, l'emplenar de brasas rojas, l'envirolar dins una coberta e anar delargar las vacas. - Es pas qu'aquò ? ço faguèt [Mowgli]. S'un dròlle o pòt far, ai pas a crentar res.

Figure 1 - Concordance trilingue serbe, anglais, occitan dans ParCoLab

Par ailleurs, l'existence de toutes ces ressources rend possible le développement d'applications permettant aux langues d'exister dans l'univers numérique, enjeu considérable pour leur vitalité : correcteurs orthographiques, claviers prédictifs, dictée vocale, reconnaissance vocale, traducteurs automatiques, outils de recherche d'information, etc. Les bases textuelles monolingues et multilingues sont également fort utiles aux enseignant(e)s et aux apprenant(e)s de ces langues.

Les défis pour outiller ces langues sont évidemment nombreux. Le premier défi est celui de réunir les données nécessaires à la construction des ressources linguistiques : celles-ci sont souvent rares, éparses ou difficilement accessibles, et rarement déjà numérisées, qu'il s'agisse des données lexicales ou textuelles brutes ou des descriptions linguistiques souvent incomplètes permettant de les analyser. Le deuxième défi est lié à la nature de ces données caractérisées par des variations graphiques et dialectales, situation fréquente pour les langues peu normées, et à la difficulté de prise en compte de la variation par les outils existants. Les autres défis, et non des moindres, sont liés à des éléments externes aux langues elles-mêmes, comme la faiblesse des moyens humains et financiers, le manque de considération fréquent pour l'entreprise d'outillage de langues perçues comme « non rentables » et le manque d'originalité en termes d'innovation souvent reproché aux quelques chercheurs courageux qui se lancent dans l'entreprise.

L'entraide et la réutilisation d'outils pour relever les défis

La mise en commun des enjeux et des défis à relever pour trois langues de France — l'alsacien, l'occitan et le picard — a permis à un petit groupe de chercheurs d'unir leurs forces pour proposer le projet RESTAURE (Ressources informatisées et Traitement Automatique pour les langues régionales), financé par l'ANR (2015-2018). Des bases textuelles étaient déjà en construction pour le picard (Picartext) et l'occitan (BaTelÒc), mais tous les autres outils et ressources étaient à créer. Le projet a permis de développer des lexiques pour l'alsacien et l'occitan, un corpus annoté en parties du discours (catégories grammaticales) pour chacune des trois langues, et de mettre en ligne une version opérationnelle de BaTelÒc. Nous n'avons pas réinventé la roue, nous avons au contraire cherché à bénéficier de l'outillage des

langues mieux dotées, pour mettre de notre côté toutes les chances de parvenir à construire des ressources de bonne qualité. Nous avons suivi l'exemple de la base textuelle du français Frantext pour nos bases textuelles ; nous avons utilisé l'étiqueteur morpho-syntaxique Tree-tagger avec son modèle pour l'allemand pour pré-annoter le corpus alsacien, et l'étiqueteur morpho-syntaxique Talismane entraîné sur le corpus catalan AnCorà-CA, pour pré-annoter le corpus occitan. Nous avons ensuite ré-annoté ces corpus manuellement, en choisissant le jeu d'étiquettes du projet Universal Dependencies (UD) conçu pour permettre l'annotation et la comparaison du maximum de langues, afin que nos ressources soient facilement réutilisables par une communauté scientifique large.

Les lexiques et les corpus collectés pour l'alsacien et l'occitan ont été aussi utilisés pour développer des claviers prédictifs pour smartphones avec le soutien de la Délégation Générale à la Langue Française et aux Langues de France. La DGLFLF nous a également soutenus pour faire entrer l'occitan dans la base de textes multilingues ParCoLab proposant aux chercheurs des textes alignés en français, serbe et anglais, mettant ainsi à disposition de la communauté des linguistes des données précieuses, voir inédites, pour comparer ces quatre langues (Figure 1).

La dynamique impulsée par ces projets a permis à l'équipe occitane de s'engager dans un projet européen avec des partenaires basques et aragonais, le projet LINGUATEC (2018-2021), financé par le programme de coopération transfrontalière de l'Union Européenne POCTEFA. Dans ce cadre, elle vient de créer le premier corpus occitan annoté en dépendances syntaxiques, toujours dans le format de Universal Dependencies (Figure 2), et en utilisant des corpus d'autres langues romanes au format UD pour initialiser l'annotation syntaxique de l'occitan. Cette expérience a largement bénéficié de l'expertise de collègues ayant réalisé le même cheminement pour outiller le serbe avec l'outil d'analyse morpho-syntaxique et syntaxique Talismane.

Le groupe de recherche informel ainsi constitué souhaite maintenant poursuivre l'outillage de l'occitan et de l'alsacien, retrouver des partenaires pour le picard et accompagner l'outillage d'autres langues de France comme le corse, le poitevin-saintongeais, etc.

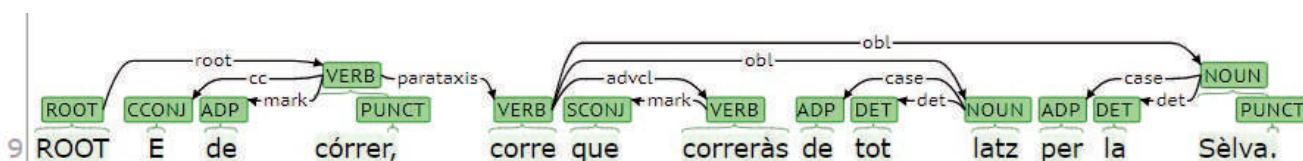


Figure 2 - Annotation en dépendances syntaxiques d'une phrase en occitan

Des expériences à leurs enseignements

Des expériences évoquées ci-dessus peuvent être tirés plusieurs enseignements.

Le premier est que l'entraide, la coopération, le partage des compétences et la mutualisation des efforts ont permis de franchir les premiers obstacles et de faire les premiers pas vers l'outillage de trois langues peu dotées, constat réconfortant en ces temps de grands isolements et de grandes extinctions, tant dans le domaine du vivant que dans celui des langues.

Le deuxième est que le retard d'outillage de ces langues peu dotées leur a finalement été favorable car elles ont pu bénéficier de l'avance des langues mieux dotées. D'une part, les analyseurs automatiques actuels « apprennent » à partir de corpus annotés, il n'est donc plus nécessaire de construire un outil spécifique pour chaque langue et il est facile d'utiliser un outil fonctionnant sur une langue proche pour effectuer une pré-annotation. On amorce ainsi le processus d'annotation, dont les phases ultérieures de correction manuelle par des équipes d'annotateurs garantissent des corpus de qualité, tout en optimisant le temps d'annotation manuelle. Ces corpus sont ensuite exploités pour entraîner les outils qui pourront à leur tour annoter automatiquement des corpus plus larges. D'autre part, des projets à visée multilingue comme *Universal Dependencies* se sont développés et facilitent grandement le développement de projets sur des langues peu ou pas du tout dotées.

Le troisième enseignement porte sur l'appréhension de la variation linguistique. La tendance en traitement automatique des langues est souvent de normaliser les données (les textes

issus des réseaux sociaux, les textes d'apprenants, etc.) pour pouvoir les traiter automatiquement avec les outils existants. La variation est finalement considérée comme un aspect négatif des données qu'on cherche à éliminer. Or, la variation est aussi une richesse car indicatrice de nombreuses informations pertinentes : origine géographique, appartenance sociale, procédés littéraires, etc. Pour outiller des langues à fortes variations, avec une vision positive de la variation, il est nécessaire de développer d'autres stratégies, qui pourront à leur tour être utiles aux recherches sur les autres langues. Il y a fort à parier que les traitements d'autres phénomènes incontournables pour les langues régionales pourront faire avancer une linguistique computationnelle au service de la description linguistique et de la diversité des langues. Nos expériences nous confortent en tous cas dans le choix de la coopération pour la mutualisation des forces et des outils, dans le choix de s'emparer de la diversité plutôt que de suivre la tendance à l'uniformisation (voir note 1), avec, en toile de fond, l'idée que la diversité linguistique reflète l'incroyable plasticité de la cognition humaine. C'est la linguistique, en tant que science humaine, qui sortira grandie d'avoir favorisé l'outillage de toutes les langues pour la description et la compréhension de cette diversité.

Myriam Bras, Dejan Stosic, Marianne Vergez-Couret, Delphine Bernard, Aleksandra Miletic, Jean Sibille

contact&info

► Myriam Bras
CLLE

myriam.bras@univ-tlse2.fr

La phylogénie des langues au service de l'analyse automatique

Récemment recruté comme chercheur CNRS au sein du laboratoire *Langues, Textes, Traitements informatiques, Cognition* (LATTICE, UMR8094, CNRS / Université Sorbonne Nouvelle / ENS Paris), Mathieu Dehouck s'intéresse à l'analyse syntaxique automatique des langues, en particulier dans une perspective multilingue. Pascal Denis est chercheur et responsable adjoint de l'équipe Inria MAGNET au sein du *Centre de Recherche en Informatique, Signal et Automatique de Lille* (CRISTAL, UMR9189, CNRS / Université de Lille / Centrale Lille). Ses recherches portent notamment sur l'apprentissage automatique pour le traitement automatique des langues, et ses applications pour l'analyse de la phrase et du discours.

Apprendre de l'histoire des langues

Les spécialistes en linguistique historique et en phylogénie essaient de reconstruire l'histoire des langues et de leur évolution en recoupant des données venant de plusieurs langues plus ou moins proches. Notre travail suit le chemin inverse, à savoir : comment l'histoire des langues, représentée sous forme d'un arbre phylogénétique, peut-elle nous informer sur les langues elles-mêmes ?

C'est une question cruciale pour le traitement automatique des langues (TAL), notamment dans la perspective de fournir des outils d'analyse pour le plus grand nombre de langues possibles, et de tenter de résorber la fracture numérique entre les populations. En effet, les quantités de données annotées disponibles pour l'apprentissage automatique de modèles d'analyse sont très inégales. Ainsi, il existe des disparités entre différentes familles de langues et, dans une même famille, certaines peuvent disposer des corpus annotés de plusieurs dizaines de milliers de phrases alors que d'autres en ont beaucoup moins, voire pas du tout.

Dans le cas précis de l'analyse syntaxique en dépendances (Figure 1), sur laquelle portent nos travaux récents, la ressource de référence est le projet *Universal Dependencies*¹, dont la [version la plus récente](#)² contient 184 corpus pour un total de 104 langues. Les annotations pour ces langues sont réparties de manière très déséquilibrée, avec une majorité de langues indo-européennes et, pour le moment, aucune langue native (d'avant la colonisation européenne) d'Amérique du nord ou du Pacifique sud, par exemple. En outre, même les langues disposant de corpus annotés présentent de grandes disparités de traitement. L'albanais, dispose ainsi d'un corpus de soixante phrases seulement pour un peu moins de 1 000 mots, alors que le tchèque compte cinq corpus totalisant plus de 127 000 phrases et 2,2 millions de mots. La conséquence de ce déséquilibre est que, à l'heure actuelle, seule une poignée de langues, les mieux dotées en corpus annotées, disposent d'analyseurs syntaxiques suffisamment précis pour donner lieu à des applications réelles.

La typologie sans la typologie

Bien que la conception d'analyseurs syntaxiques procèdent le plus généralement de façon indépendante pour chaque langue, il existe néanmoins déjà quelques approches, dites par transfert, pour créer des modèles pour des langues peu ou non dotées à partir des données de langues bien dotées plus ou moins proches. En revanche, très peu d'approches exploitent l'histoire des langues comme source d'information.

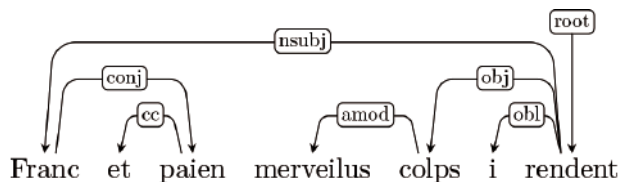


Figure 1 - Un exemple de phrase annotée avec un arbre syntaxique en dépendances. Cette phrase est extraite du corpus Ancien Français SRCMF disponible dans le projet Universal Dependencies (Petrov S. and Stejnka A. (eds.) 2013, *Syntactic Reference Corpus of Medieval French* (SRCMF), ENS de Lyon, IIR Stuttgart).

À titre d'exemples, Naseem et al.³ et Aufrant et al.⁴ se sont attaqués au problème de l'apprentissage automatique d'analyseurs syntaxiques en utilisant de l'information typologique (ordre des mots, existence d'articles, existence d'un verbe copule) pour guider le partage de l'information entre les différentes parties de leurs modèles. Ces approches souffrent d'une limitation importante, à savoir : la nécessité d'avoir accès à des informations typologiques fiables sur les langues concernées, qui sont rarement disponibles pour les langues peu dotées.

Pourtant, les arbres phylogénétiques (Figure 2) encodent également de l'information typologique, mais de manière implicite. En effet, bien que les reconstructions phylogénétiques se basent le plus souvent sur des mesures lexicales, souvent le seul matériau disponible, l'on s'attend néanmoins à ce que des langues historiquement proches partagent aussi des traits grammaticaux. Ce que nous procurent les arbres phylogénétiques est donc une notion de similarité entre les langues, sans nécessairement connaître la nature exacte des traits grammaticaux qu'elles partagent. Ceux-ci pourront néanmoins être inférés par les méthodes d'apprentissage à partir des données annotées. En outre, un avantage supplémentaire des arbres phylogénétiques par rapport aux traits typologiques est qu'il existe plus de méthodes pour créer des arbres phylogénétiques automatiquement à partir de listes de mots que pour trouver des traits typologiques à partir de texte.

Apprendre des modèles de manière évolutive

Notre approche rompt avec les travaux précédents, dans la mesure où nous proposons d'apprendre simultanément les modèles d'analyse associés aux différentes langues en exploitant directement l'arbre phylogénétique sous-jacent, et sans autre information typologique.

Les modèles d'analyse syntaxique, appris automatiquement sur des corpus annotés, sont des représentations très abstraites et sommaires de la grammaire des langues, mais l'on peut supposer

1. Zeman D., Nivre J., Abrams M. et al. 2020, *Universal dependencies 2.7*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

2. Voir la [version 2.7](#), parue en novembre 2020.

3. Naseem T., Barzilay R. et Globerson A. 2012, Selective sharing for multi-lingual dependency parsing, in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers* - Volume 1: 629–637, Stroudsburg, PA, USA.

4. Aufrant L., Wisniewski G. et Yvon F. 2016, Zero-resource Dependency Parsing: Boosting Delexicalized Cross-lingual Transfer with Linguistic Knowledge, in *Proceedings of the 26th International Conference on Computational Linguistics*, pp119–130, Osaka, Japan.

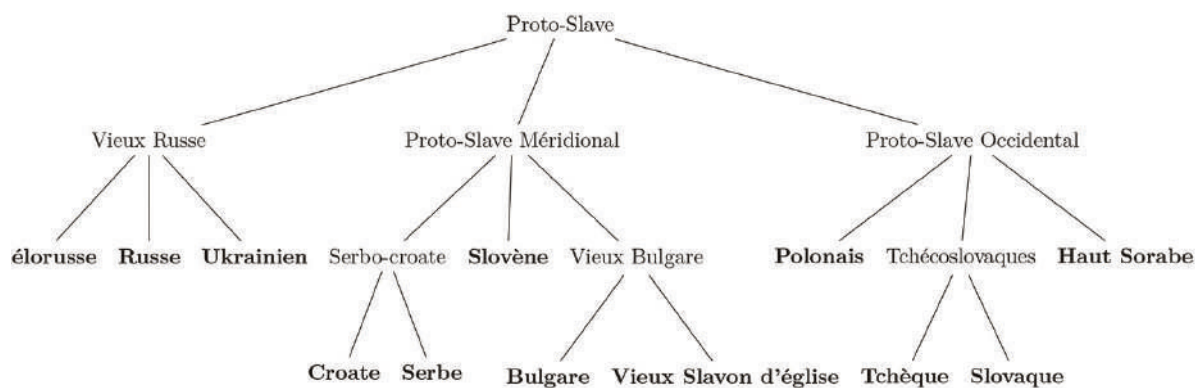


Figure 2 - Exemple d'arbre phylogénétique. Rameau slave de l'arbre phylogénétique utilisé pour nos travaux. Les noms des nœuds intermédiaires sont indicatifs, seules la structure de l'arbre et la place des feuilles (en gras) sont importantes. Le pluriel tchécoslovaque réfère aux langues du groupe tchécoslovaque et non pas à la norme écrite utilisée dans les années 1920 et 1930.

qu'ils évoluent d'une manière similaire à ces langues. Par exemple, si l'on avait des corpus pour les langues romanes représentant leur état tous les cinquante ans en commençant en l'an zéro et jusqu'aujourd'hui, l'on pourrait s'attendre à ce que des modèles d'analyse, appris sur chacun de ces corpus, forment une structure similaire à celle proposée par les linguistes historiques pour les langues romanes elles-mêmes. L'on s'attendrait entre autres à ce que des modèles représentant des corpus proches en terme de dates et de variétés (picard de 1150 et normand de 1300), soient plus proches que des modèles représentant des variétés très différentes (portugais de 1900 et roumain de 1300).

L'on pourrait alors faire évoluer les modèles lentement pour suivre l'évolution des langues au plus près, gardant les traits hérités d'époques antérieures et remplaçant les archaïsmes par leur formes modernes.

Malheureusement, nous n'avons pas accès à ce genre de données, entre autres parce que *Universal Dependencies* ne dispose très majoritairement d'annotations que pour des langues encore vivantes, beaucoup de langues n'ayant d'ailleurs commencé à être écrites que très récemment. Cependant, l'on peut utiliser un mélange de données venant de langues apparentées pour simuler leur langue mère. Plus précisément, l'on considèrera comme exemple de la langue mère l'ensemble des annotations associées à ces descendants. Bien que cette approche soit très simpliste et qu'elle ait peu de chance de représenter fidèlement la grammaire de la langue mère, c'est peut-être là une de ses forces. En effet, l'objectif final étant d'apprendre des modèles d'analyse pour les langues modernes⁵, il paraît pertinent de confronter les modèles à des données similaires à celles qu'ils auront à traiter à la fin et ce, dès le début de l'apprentissage.

Nous proposons alors de faire « évoluer » des modèles d'analyse en parallèle pour plusieurs langues historiquement apparentées (Figure 3). L'idée est qu'en entraînant un modèle d'analyse sur des données venant de plusieurs langues proches, on l'incite à se concentrer sur les caractéristiques communes à ces langues, dans notre cas, sur les traits syntaxiques partagés. L'on peut ensuite faire des copies du modèle appris pour imiter la langue mère et les passer aux rameaux descendants pour continuer l'apprentissage en partant d'un modèle déjà plus robuste. En s'enfonçant de plus en plus profondément dans l'arbre, les modèles se spécialisent de

plus en plus. Appris pour une famille de langues au départ, ils se spécialisent ensuite sur un rameau et, enfin, sur une seule langue comme représenté dans la figure 2. Cependant, comme ils ont vu des données d'autres langues plus ou moins proches de leur langue cible, ils ont emmagasiné de l'information syntaxique qui, tout en étant utile à l'analyse de la langue cible, n'aurait pas nécessairement été disponible dans le seul corpus de la langue cible.

Résultats empiriques

Nous avons mené des expériences sur un ensemble de corpus annotés avec des arbres de dépendances rendus disponibles par le projet *Universal Dependencies*. Nous avons entraîné des modèles à reconstruire les arbres en dépendances à partir de mots de la phrase et de leurs attributs morphologiques (genre, personne, cas, temps, mode, etc.). Les modèles étaient soit entraînés ensemble le long d'un arbre phylogénétique représentant les relations des différentes familles et sous familles de langues, soit entraînés indépendamment directement pour chacune des langues disponibles.

Les modèles (linéaires et neuraux) entraînés dans un arbre phylogénétique représentant une partie de l'histoire évolutive des langues ont montré de meilleurs performances que des modèles entraînés avec les mêmes données mais de manière indépendante pour chaque langue. Les performances des modèles phylogénétiques sont d'autant plus remarquables que les langues en question sont peu dotées et que leur famille linguistique est bien fournie en terme de langues.

Un exemple de ce phénomène est donné par le haut sorabe, une langue slave parlée dans l'est de l'Allemagne par environ 15 000 personnes. Bien qu'elle soit très peu dotée (une trentaine de phrases d'entraînement chez *Universal Dependencies*), sa proximité avec le polonais, le tchèque et le slovaque fait que l'on peut entraîner des modèles fiables en se basant sur les données disponibles pour ses langues sœurs.

En outre, les modèles entraînés pour un ensemble de langues proches peuvent aussi servir de point de départ pour analyser des langues pour lesquelles nous n'avons pas de données d'apprentissage du tout.

5. Par « langues modernes », nous entendons des langues pour lesquelles nous avons des données. En ce sens, le latin, le gotique ou encore le sanskrit sont des langues modernes, car il existe une littérature dans ses langues qu'il est intéressant d'analyser avec des outils numériques modernes.

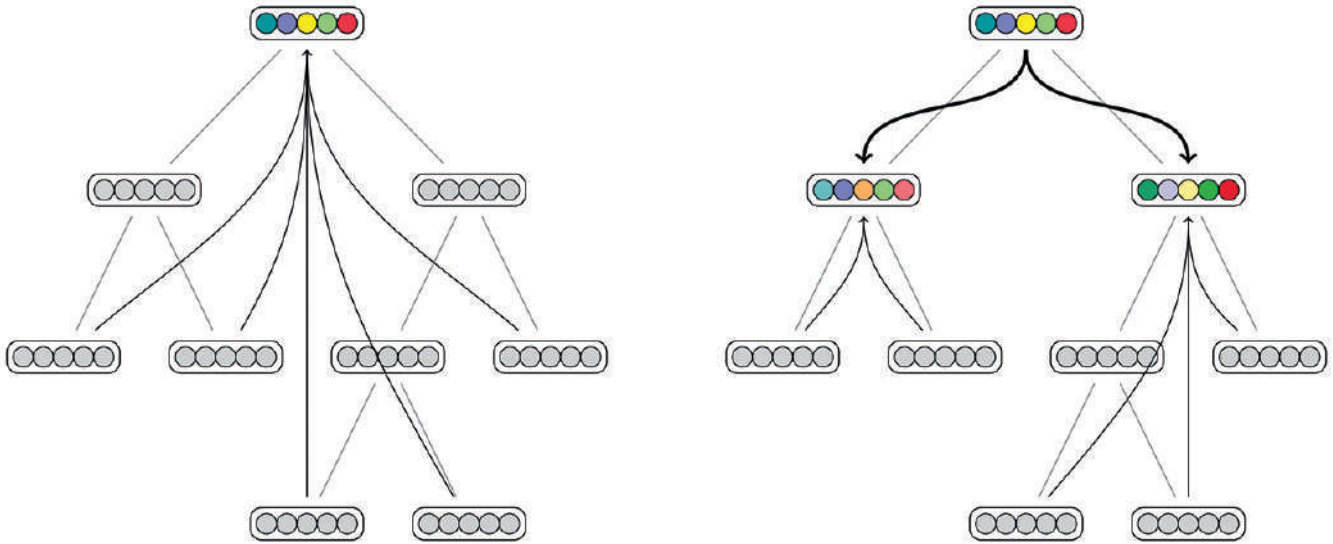


Figure 3. Représentation de la méthode d'apprentissage phylogénétique. Les données de toutes les feuilles sont utilisées pour entraîner le modèle de la racine (gauche). Le modèle est ensuite copié et passé à ses descendants directs où l'apprentissage reprend (droite). Chaque nœud n'a accès qu'aux données de ses propres descendants.

Conclusion

Nos travaux montrent que les modèles utilisés pour le TAL, en l'occurrence des modèles d'analyse syntaxique, peuvent tirer profit de l'information historique sur l'évolution des langues, représentée ici par un arbre phylogénétique, pour apprendre mieux des données disponibles en partageant les ressources existantes pour différentes langues.

Une des limitations de l'approche présentée ici est qu'elle utilise un arbre phylogénétique comme source d'information historique. Or, les arbres phylogénétiques ne sont pas adaptés à la représentation des transferts horizontaux⁶ que l'on peut observer dans les aires linguistiques, par exemple. Les traits acquis par ces biais sont toutefois présents dans les langues et nous devons les y analyser. Il nous faudra donc envisager des structures plus expressives que les arbres, comme les graphes dirigés acycliques⁷, si nous voulons avoir des représentations plus fidèles de l'évolution des langues. De

plus, il nous faudra des méthodes pour créer automatiquement ces nouvelles structures à partir de données textuelles si nous voulons les utiliser à plus grande échelle et pour les langues sous dotées.

Nous invitons le lecteur intéressé à se référer à Dehouck et Denis⁸ pour plus de détails concernant les algorithmes utilisés et pour des résultats complets.

contact&info

▶ Mathieu Dehouck
Lattice
mathieubmddehouck@mailoo.org

▶ Pascal Denis
CRISTAL
pascal.denis@inria.fr

6. Transferts horizontaux : emprunts de mots ou de structures grammaticales à des langues par contact, sans forcément que celles-ci ne partagent un ancêtre commun.

7. Graphes dirigés acycliques : structure dans laquelle les nœuds peuvent avoir plusieurs parents directs. Dans un arbre, les nœuds n'ont qu'un parent.

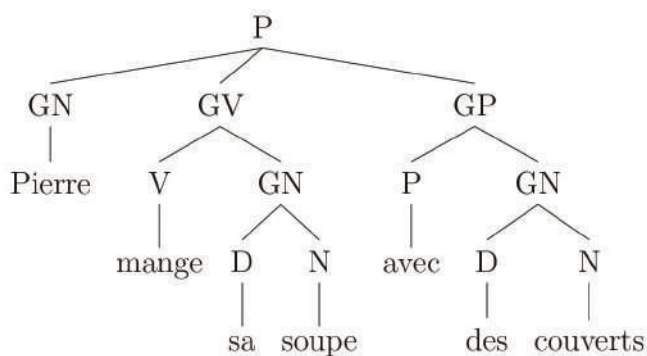
8. Dehouck M. et Denis P. 2019, Phylogenetic multi-lingual dependency parsing, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers), pp 192–203, Minneapolis, Minnesota.

Analyse syntaxique automatique

Professeur à l'Université de Paris, membre honoraire de l'institut universitaire de France (IUF), Benoit Crabbé est affilié au [Laboratoire de Linguistique Formelle \(LLF, UMR7110, CNRS / Université de Paris\)](#). Ses recherches portent particulièrement sur la linguistique informatique et la compréhension automatique du langage avec un intérêt particulier pour l'analyse syntaxique du français et des langues apparentées.

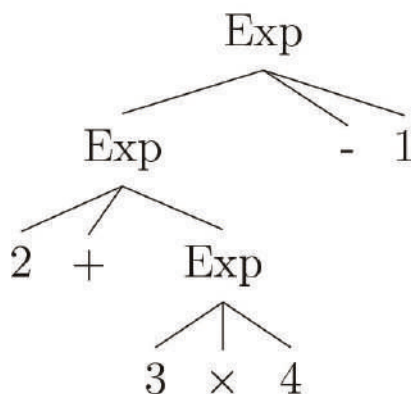
Cet article présente dans les grandes lignes l'évolution des modèles d'analyse syntaxique automatique depuis les origines théoriques en linguistique jusqu'aux modèles construits par apprentissage statistique à partir de grandes masses de données textuelles.

En linguistique informatique, l'analyse syntaxique automatique consiste à prédire automatiquement la structure de la phrase à l'aide d'un arbre, comme par exemple :



Un arbre d'analyse syntaxique qui représente une phrase (P) constituée de groupes nominaux (GN), verbaux (GV) et prépositionnels (GP). Les catégories des mots représentent les noms (N), les déterminants (D), les prépositions (P) et les verbes (V)

L'analyse syntaxique pour les langues naturelles partage des propriétés essentielles avec les modèles utilisés pour structurer les langages artificiels. Par exemple, il est habituel de structurer une expression arithmétique par un arbre de manière analogue à une phrase en langue naturelle :



Un arbre d'analyse pour une expression arithmétique (Exp) dont les sous-expressions sont également des expressions arithmétiques

L'analyse syntaxique est vue comme une étape préliminaire à l'évaluation sémantique d'une phrase ou d'un énoncé. Pour le cas des expressions arithmétiques, l'arbre peut servir à calculer compositionnellement sa valeur de manière ascendante. Sur cet exemple, on calculera que la sous-expression '3 x 4' a pour valeur 12. Cette valeur permet à son tour d'évaluer l'expression '2+(3 x 4)' qui aura pour valeur 14, et l'on terminera par une évaluation de l'expression dans son ensemble '2+(3 x 4)-1' qui a pour valeur 13.

Sur les aspects syntaxiques, l'analyse des langues naturelles se distingue toutefois des langages formels par des effets apparents de mouvement. Dans la langue des expressions arithmétiques, les foncteurs ou opérateurs (+ - x /) sont situés au même niveau de l'arbre que leurs arguments, ce qui permet d'exprimer relativement aisément une méthode de calcul du sens. En revanche, en langue naturelle, cette propriété n'est pas vérifiée. Certains mots semblent déplacés. Ainsi, pour construire une phrase interrogative en français, l'élément interrogé est comme déplacé depuis sa position ordinaire jusqu'à la position interrogée. En voici une illustration :

- 1) Tu crois que Marie a vu le loup hier
- 2) Quel loup crois-tu que Marie a vu hier ?

Le complément d'objet direct *quel loup* est comme déplacé depuis sa position ordinaire, adjacente au verbe, vers une position qui marque l'interrogation.

Les modèles symboliques

En linguistique, les premiers modèles de Noam Chomsky¹ sont structurés en deux composantes : une composante générative qui engendre des phrases canoniques, comme dans la phrase 1 ci-dessus, dans lesquelles l'évaluation sémantique est facile : arguments et foncteurs sont situés au même niveau de l'arbre. Une composante transformationnelle est chargée de déplacer certains éléments vers une position de surface comme dans le cas des phrases interrogatives.

Avec l'apparition des ordinateurs et l'essor de l'informatique, il est apparu intéressant de simuler numériquement les différents modèles de la grammaire transformationnelle : écrire la grammaire d'une langue, c'est-à-dire l'ensemble des règles qui caractérisent la structure des phrases bien formées et les transformations qui leur sont associées est une tâche non triviale dès qu'on cherche à atteindre un certain niveau de précision : des problèmes de cohérence de la grammaire et de complétude de l'ensemble de règles surgissent inévitablement. La simulation informatique peut alors apporter des méthodes qui permettent de tester la grammaire et de traiter ces deux types de problèmes.

1. Chomsky N. 1956, Three models for the description of language, IRE Transactions on Information Theory 2 (3) ; Chomsky N. 1957, *Syntactic Structures*, Mouton, The Hague.

Toutefois, il est rapidement apparu que les premiers modèles étaient indécidables², c'est-à-dire qu'il n'est pas possible de concevoir un algorithme capable de dire à coup sûr et en un temps fini, pour toute suite de mots, s'il est possible d'en construire son arbre d'analyse. C'est la composante transformationnelle qui est responsable de l'indécidabilité.

Sur les aspects syntaxiques, la linguistique computationnelle s'est développée en cherchant à définir des grammaires qui expriment la syntaxe des langues naturelles sans transformations. Il s'agit du paradigme des grammaires d'unification³. La motivation première a été de réexprimer les grammaires de langue naturelle en utilisant des méthodes utilisées par ailleurs en informatique pour l'analyse des langages de programmation. Il s'agit, d'une part, d'exprimer les grammaires avec un nombre raisonnable de règles grâce à un mécanisme appelé unification et, d'autre part, d'obtenir une évaluation sémantique des phrases en utilisant des méthodes inspirées à la fois des travaux de Montague⁴ et des développements en sémantique des langages de programmation.

Le tournant statistique

Les nouveaux modèles grammaticaux ont permis d'écrire des grammaires informatisées couvrant des fragments importants de la langue naturelle. Toutefois, ceux-ci se sont heurtés à un obstacle important : celui de l'ambiguïté. Considérons l'exemple suivant :

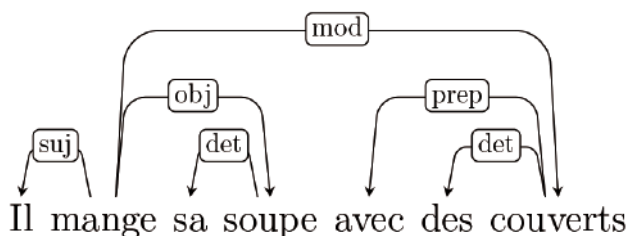
- 3) Pierre mange une salade avec des couverts
- 4) Pierre mange une salade avec des champignons

On remarque que ces deux phrases ont strictement la même forme. En revanche, elles diffèrent sur un élément de vocabulaire. Cette différence entraîne un changement quant à la structuration et à l'interprétation de la phrase : dans la phrase 3, *couvert* est un instrument pour *manger* alors que dans la phrase 4, *champignons* est un ingrédient de la *salade*.

Pour un humain, ces exemples ne posent en principe pas problème : il n'est pas naturel d'interpréter *champignons* comme un instrument de *manger* ou *couverts* comme un ingrédient de *salade*. On dispose de connaissances lexicales et encyclopédiques implicites qui permettent de déterminer l'interprétation. Mais, pour un ordinateur, la totalité de cette connaissance est difficilement accessible et les différentes interprétations sont en principe possibles.

Face à ce problème d'ambiguïté, les grammaires écrites manuellement ne sont pas équipées d'un mécanisme naturel qui permet de décider de l'interprétation à préférer. Dans le cas des grammaires couvrant de très larges fragments de la langue, les analyses possibles se comptent souvent en dizaine de milliers à tel point que l'ensemble des interprétations devient inutilisable.

Pour traiter les problèmes d'ambiguïté, les modèles d'analyse ont été progressivement augmentés d'une composante statistique. À cette fin, de larges corpus de textes ont été annotés en arbres. Intuitivement, les modèles d'analyse statistique sont alors entraînés pour produire des analyses d'autant plus probables qu'elles ressemblent aux arbres du corpus sur lequel les probabilités du modèle ont été estimées. L'application naïve de méthodes statistiques est toutefois limitée. Certes, il est possible d'apprendre facilement des modèles privilégiant les analyses qui comportent des structures grammaticales courantes.



Arbre de dépendances où les relations entre les mots sont illustrées par des flèches. La relation sujet est notée (suj), la relation objet (obj), la relation de modification (mod), la relation de détermination (det) et la relation de préposition (prep).

Le problème de désambiguïté concerne principalement les problèmes de relations lexicales illustrées en phrases 3 et 4, ainsi que les problèmes de portée de la coordination. Il est ainsi crucial de modéliser les relations entre mots comme entre *manger* et *couverts* ou entre *salade* et *champignons*⁵. C'est ainsi que le tournant statistique a été accompagné d'un retour des grammaires de dépendances⁶, formalisées préalablement notamment par Lucien Tesnières⁷. Celles-ci ont pour propriété de représenter la phrase à l'aide de relations entre mots. La popularité actuelle des grammaires de dépendances s'explique aussi par une meilleure capacité à représenter les langues à ordre des mots libres ce qui a entraîné une évolution des perspectives vers l'analyse automatique multilingue : le projet *Universal Dependencies*, qui regroupe des corpus arborés pour plus de cent langues différentes, est emblématique de cette dynamique.

L'apprentissage profond

Malgré leur taille déjà considérable, les corpus annotés manuellement en arbres restent limités pour apprendre des modèles statistiques de désambiguïté. À titre indicatif, le *Penn Treebank*⁸ comporte à peu près 50 000 phrases pour sa partie journalistique et le *French treebank*⁹ près de 20 000 phrases. Vu qu'une partie importante du problème de désambiguïté demande des relations entre couples de mots, ces corpus, si gros soient-ils en termes d'annotation manuelle, restent de taille dérisoire pour acquérir les statistiques biléxicales.

2. Peters S., Ritchie R. 1973, On the generative power of transformational grammars, *Information Sciences*, 6.
 3. Bresnan J. 1982, *The mental representation of grammatical relations*, MIT Press ; Gazdar G., Klein E., Pullum G., Sag I. 1985, *Generalized Phrase Structure Grammar*, Harvard University Press ; Pollard C., Sag I. 1994, *Head Driven Phrase Structure Grammar*, University of Chicago Press.
 4. Montague R. 1974, *Formal Philosophy. Selected Papers of R. Montague*, ed. by R. Thomason, Yale University Press.
 5. Collins M. 2003, Head-Driven Statistical Models for Natural Language Parsing, *Computational Linguistics*, 29(4).
 6. Nivre J., Scholz M. 2004, *Deterministic Dependency Parsing of English Text*, COLING ; McDonald R., Pereira F., Ribarov K., Hajič J. 2005, *Non-Projective Dependency Parsing using Spanning Tree Algorithms*, HLT-EMNLP.
 7. Tesnières L., 1959, *Éléments de syntaxe structurale*, Klincksieck.
 8. Marcus M., Santorini B., Marcinkiewicz M. 1993, Building a Large Annotated Corpus of English: The Penn Treebank, *Computational Linguistics* 19(2).
 9. Abeillé, A., Clément L., Toussenet F. 2003, Building a treebank for French, in A. Abeillé (ed.) *Treebanks*, Kluwer, Dordrecht.

Les méthodes d'analyse fondées sur l'apprentissage profond sont parvenues dans une large mesure à faire sauter le verrou de l'annotation en construisant des modèles qui parviennent à extraire de manière non supervisée des régularités statistiques issues de texte brut en s'appuyant sur des vecteurs. Une première avancée a eu lieu avec les modèles de la famille *word2vec*¹⁰ qui permet de représenter des mots sémantiquement similaires par des vecteurs similaires. Une seconde amélioration provient de modèles de la famille *transformers*¹¹ dont le représentant le plus connu est appelé BERT¹². Ce dernier modèle permet également de spécialiser/désambiguïser les vecteurs de mots en fonction du contexte de la phrase.

La combinaison de meilleures représentations lexicales issues de l'arrivée de l'apprentissage profond avec les algorithmes traditionnels produit des analyseurs syntaxiques à l'état de l'art.

Ceux-ci donnent des résultats très fiables et comparables à ceux d'experts humains pour certains types de textes. L'un des enjeux scientifiques des prochaines années sera certainement de se donner les moyens de mieux contrôler les modèles par apprentissage profond, pour permettre de faciliter leur interprétation et affiner les hypothèses de travail, notamment dans le but de contraster ces modèles avec ceux de la tradition scientifique en linguistique et en psycholinguistique.

contact&info

► Benoît Crabbé
LLF

benoit.crabbe@linguist.univ-
paris-diderot.fr

10. Mikolov T. et al. 2013, Efficient Estimation of Word Representations in Vector Space, *International Conference on Learning Representations*.

11. Vaswani A. et al. 2017, Attention Is All You Need, *Neural Information Processing Systems*.

12. Devlin J. et al 2019, « BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *North American Chapter of the Association for Computational Linguistics*.

Outils computationnels pour l'étude des mécanismes d'adaptation à la variation en perception de la parole

Olivier Crouzet est enseignant-chercheur en sciences du langage à l'Université de Nantes et membre du *Laboratoire de Linguistique de Nantes (LLING, UMR6310, CNRS / Université de Nantes)*. Ses activités de recherche portent notamment sur l'étude des mécanismes de perception et de production de la parole dans la population adulte en général et chez les personnes souffrant de handicaps sensoriels ou moteurs. Il s'intéresse à la modélisation du type de représentations et de mécanismes mentaux qu'on utilise lorsqu'on parle et/ou écoute la parole.

La parole humaine est un phénomène acoustique qui présente une grande variabilité¹, laquelle se manifeste à travers les nombreuses formes physiques possibles qu'un locuteur peut utiliser pour produire une catégorie sonore. Ainsi, pour n'en donner qu'un exemple parmi tant d'autres, la production de la voyelle /i/ peut correspondre à un très grand nombre de réalisations différentes². Cette variation physique n'est pas le seul phénomène qui sous-tend la complexité des mécanismes de perception : les auditeurs s'adaptent également au contexte du son traité. Ainsi, l'identification d'un son est influencée par les sons environnants. Ces effets sont à la fois liés à des mécanismes d'adaptation au contexte acoustique³ et à l'utilisation d'informations de haut niveau comme le sens d'une phrase⁴ ou l'existence d'un mot dans la langue⁵.

Afin de mieux comprendre l'impact de cette variabilité sur la perception ainsi que les mécanismes cognitifs mis en œuvre par les auditeurs pour s'adapter à cette variation, l'étude de la perception de la parole par l'humain a eu recours, dès les années 1950, aux outils technologiques puis informatiques qui commençaient à voir le jour. Nous nous limiterons ici à deux classes d'outils computationnels qui peuvent être intéressants dans ce domaine : d'une part, des outils qui permettent de manipuler les propriétés acoustiques de sons de parole ; d'autre part, des méthodes qui permettent de quantifier certaines propriétés du sens des phrases.

Concernant la manipulation des propriétés acoustiques, l'un des principaux outils utilisés par les chercheurs dans ce domaine a été la synthèse vocale paramétrique, dont le précurseur à la fin des années 1940 fut un instrument analogique appelé le *pattern playback*. Cet outil avait été conçu par Franklin S. Cooper et son équipe aux *Laboratoires Haskins* pour transformer l'image d'un spectrogramme (une représentation temps-fréquence des signaux sonores comme on peut en voir sur la figure 1) en son. Il devenait ainsi possible de dessiner ou de modifier à la main l'image d'un spectrogramme et d'obtenir le résultat acoustique de ces modifications. On pouvait alors faire écouter ces signaux sonores à des auditeurs humains pour étudier les effets des transformations effectuées. Franklin S. Cooper et son équipe⁶ ont ainsi eu recours au *pattern playback* pour générer de fines variations des fréquences de résonance du conduit vocal similaires à celles qu'on peut voir sur la figure 1. Plus tard, des outils informatiques ont été implémentés en s'inspirant du *pattern playback*⁷. La synthèse vocale dite « paramétrique » offrit alors la possibilité de générer des signaux de parole de manière totalement synthétique sur un ordinateur (Figure 1) en contrôlant très précisément les paramètres acoustiques souhaités.

Donald Broadbent et son équipe⁸ ont utilisé cette technique pour étudier les effets de la variation acoustique « contextuelle » sur la perception. Les participants devaient répéter un mot cible (en anglais : *bit, bet, bat* ou *but*) précédé d'une phrase de consigne

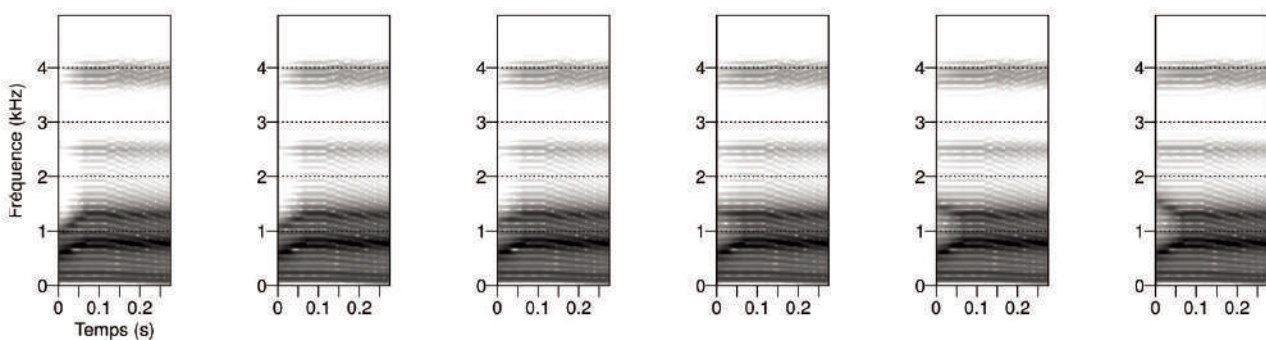


Figure 1 - Des variations fines de la transition associée à la deuxième résonance du conduit vocal sont ici réalisées à l'aide d'une synthèse paramétrique de Klatt. Elles conduisent à générer un continuum acoustique qui passe « progressivement » d'un /ba/ à un /da/.

1. Meunier C. 2005, Invariants et variabilité en phonétique, in Nguyen N., Wauquier-Gravelines S. & Durand J. (Éds.), *Phonologie et Phonétique : Forme et Substance*, chapitre 13 : 349–374. Lavoisier.
2. Peterson G. E. & Barney H. L. 1952, *Control Methods Used in a Study of the Vowels*, *The Journal of the Acoustical Society of America*, 24(2) : 175–184.
3. Ladefoged P. & Broadbent D. E. 1957, *Information conveyed by vowels*, *The Journal of the acoustical society of America*, 29(1) : 98–104 ; Mann V. 1980, Influence of preceding liquid on stop-consonant perception, *Perception & Psychophysics*, 28 : 407–412.
4. Miller J. L., Green K. & Schermer T. M. 1984, *A distinction between the effects of sentential speaking rate and semantic congruity on word identification*, *Perception & Psychophysics*, 36(4) : 329–337.
5. Ganong W. F., 1980, *Phonetic categorization in auditory word recognition*, *Journal of Experimental Psychology : Human Perception & Performance*, 6 : 110–125.
6. Cooper F. S., Delattre P. C., Liberman A. M., Borst J. M. & Gerstman L. J. 1952, *Some Experiments on the Perception of Synthetic Speech Sounds*, *Journal of the Acoustical Society of America*, 24(6) : 597–606.
7. Klatt D. 1980, *Software for a cascade / parallel formant synthesizer*, *The Journal of the Acoustical Society of America*, 67 : 971–995.
8. Broadbent D. E., Ladefoged P. & Lawrence W. 1956, *Vowel sounds and perceptual constancy*, *Nature*, 178(4537) : 815–816.

« *Please say what this word is...* ». Cette portion de phrase était altérée au niveau des fréquences de résonance du conduit vocal afin de créer des « contextes acoustiques » différents sans altérer l'identification des mots qui la composaient (imaginez, par exemple, que vous parliez en ayant tendance à serrer vos mâchoires : les propriétés acoustiques des sons produits changeraient sans conduire à entendre des mots différents). Les réponses perceptives des participants étaient toujours réalisées sur une cible acoustique non modifiée. Néanmoins, les auteurs montrent que les réponses perceptives sont systématiquement influencées par le contexte acoustique de la phrase : lorsqu'on transforme acoustiquement la phrase de consigne, *bit* peut être perçu comme *bet* ou l'inverse. Ces changements de catégorisation perceptive sont interprétés comme des effets de compensation : si la cible ne change pas, les variations de son environnement modifient sa perception. Peter Ladefoged⁹ a ultérieurement répliqué ces résultats avec des signaux produits naturellement. Ces outils synthétiques contribuent cependant toujours à la recherche actuelle par leur capacité à donner accès à des manipulations expérimentales fines de la parole. Ces phénomènes ont par exemple été récemment approfondis par Matthias J. Sjerps et son équipe¹⁰.

Plus récemment, de nouveaux outils sont venus compléter la synthèse paramétrique en donnant accès à la possibilité de manipuler très précisément des signaux de parole naturelle. En effet, l'une des limites de la synthèse paramétrique est qu'elle produit des signaux qui sont parfois trop éloignés de signaux naturels. Les outils commerciaux actuels de synthèse vocale sont capables de produire des formes beaucoup plus naturelles mais ne permettent pas toujours aux chercheurs de manipuler les phénomènes qui les intéressent. Hideki Kawahara et son équipe¹¹ ont conçu un ensemble d'algorithmes qui permettent aujourd'hui de manipuler très précisément certains paramètres de signaux de parole naturels. Cet outil est développé sous la forme d'une boîte à outils *Matlab / Octave* et une version est mise à disposition sous licence libre. Il est fondé sur un modèle source-filtre de la production de la parole et permet par exemple, à partir d'un enregistrement réel, de manipuler des propriétés acoustiques liées soit à la source sonore (fréquence fondamentale de vibration des cordes vocales et composition spectrale), soit au filtre (longueur du conduit vocal, fréquences de résonance du conduit vocal). Il ne remplace pas la synthèse paramétrique mais il donne accès à des manipulations de signaux naturels qui n'étaient pas possibles auparavant : on peut par exemple générer une suite de stimuli allant progressivement d'un signal naturel à un autre (en anglais, « *morphing* » / en français, « morphose »). Il est aussi possible de partir d'un signal naturel et de transformer des paramètres comme la hauteur de la voix ou les valeurs des fréquences de

résonance du conduit vocal. Cet outil a, par exemple, été utilisé pour étudier les contributions respectives de la hauteur de la voix et de la longueur du conduit vocal pour la perception du genre des locuteurs¹².

Nous avons évoqué précédemment l'influence que peuvent jouer les informations provenant du sens de la phrase sur la catégorisation phonétique. Ces phénomènes ont été initialement mis en évidence par Joanne L. Miller et son équipe, avec un matériel constitué de paires minimales de mots Consonne-Voyelle-Consonne où la variation portait sur le caractère voisé / non-voisé de la consonne initiale (par exemple, en anglais : *bath / path*, les deux mots s'opposent sur la base du « voisement » qui correspond à la présence — voisé — vs l'absence — non-voisé — de vibration des cordes vocales pendant une phase précise de la production). Ces paires avaient subi un traitement pour créer des stimuli intermédiaires entre les deux extrêmes. La cible ambiguë était précédée de deux phrases introductrices possibles, chacune tendant vers le sens d'un seul mot de la paire (« *She needs hot water for the _* » vs « *She likes to jog along the _* »). Comme l'indiquent les auteurs, l'influence sémantique de la phrase avait été évaluée de manière intuitive. Plus récemment, afin de mieux contrôler ces jugements intuitifs, on a mis en place des expériences pilotes dans lesquelles on s'assure que les auditeurs interprètent les stimuli ambigus de manière contextuellement appropriée¹³.

Aujourd'hui, les méthodes d'exploration automatisée de corpus langagiers permettent de produire des mesures quantitatives des relations sémantiques entre les mots d'une phrase. À partir de l'exploration des relations de proximité physique entre mots dans un grand corpus, on peut représenter chaque mot comme un vecteur multi-dimensionnel et quantifier la proximité sémantique entre deux mots à partir d'une mesure d'angle entre les deux vecteurs qui les caractérisent. Ces mesures dites de plongements de mots (*word embeddings / word2vec*) peuvent avantageusement compléter, voire remplacer, les jugements intuitifs et reposent sur une représentation distribuée des mots dans un espace vectoriel¹⁴. Ces auteurs ont développé des méthodes qui optimisent l'apprentissage de ces représentations vectorielles sur la base de données textuelles non-structurées de très grande taille. De tels outils permettent d'estimer la similarité sémantique entre deux mots sur une échelle de valeurs allant de -1 (sens opposé) à 1 (similarité maximale de sens) en passant par 0 (aucun lien sémantique). Des modèles de plongements de mots fondés sur des corpus de langue française sont librement disponibles¹⁵ et peuvent être mis en œuvre avec la bibliothèque *GenSim* sous Python.

9. Ladefoged P. 1989, *A note on "Information conveyed by vowels"*, *The Journal of the Acoustical Society of America*, 85(5) : 2223–2224.

10. Sjerps M. J., McQueen J. M. & Mitterer H. 2013, *Evidence for precategory extrinsic vowel normalization*, *Attention, Perception, & Psychophysics*, 75(3) : 576–587.

11. Kawahara H., Morise M., Takahashi T., Nisimura R., Irino T. & Banno H. 2008, *Tandem-STRAIGHT : A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation*, in 2008 *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp 3933–3936 : IEEE.

12. Skuk V. G. & Schweinberger S. R., 2014, *Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender*, *Journal of Speech, Language, and Hearing Research*, 57(1) : 285–296 ; Fuller C. D., Gaudrain E., Clarke J. N., Galvin J. J., Fu Q.-J., Free R. H. & Ba Skent D. 2014, *Gender Categorization Is Abnormal in Cochlear Implant Users*, *Journal of the Association for Research in Otolaryngology*, 15(6) : 1037–1048.

13. Gow D. W. & Olson B. B. 2016, *Sentential influences on acoustic-phonetic processing : a Granger causality analysis of multimodal imaging data*, *Language, Cognition and Neuroscience*, 31(7) : 841–855.

14. Mikolov T., Chen K., Corrado G. S. & Dean J. 2013, *Efficient estimation of word representations in vector space*.

15. Fauconnier J.-P. 2015, French word embeddings. Gaudrain E. & Crouzet O. 2019, *word2vec model trained on lemmatized French Wikipedia 2018*, Rapport interne, Funding : CNRS International Mobility Support program, 2017 ; European Community PRESTIGE Mobility program No 2017–2–0044 ; NWO / ZonMW VICI : 918–17–603 ; LABEX CeLyA (ANR–10–LABX–0060) of Université de Lyon, within the program « Investissements d'Avenir » (ANR–16–IDEX–0005) operated by the French National Research Agency (ANR).

Contexte sémantique

- Contexte 0
- Contexte 1
- Contexte 2

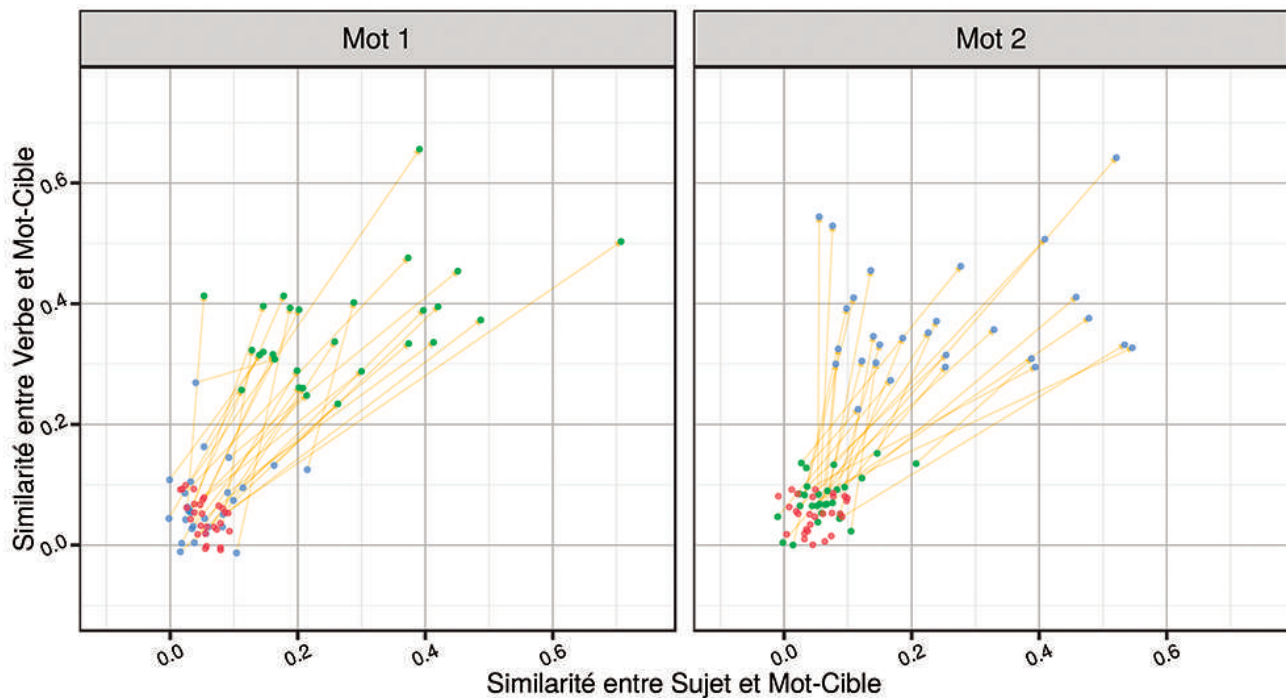


Figure 2 - Valeurs de similarité entre le mot-cible et respectivement le sujet (abscisse) / le verbe (ordonnée). Chaque point correspond à un mot d'une paire (Mot-cible 1, par exemple « balle », dans le graphique de gauche vs mot-cible 2, par exemple « belle », dans le graphique de droite). Les points verts correspondent aux phrases porteuses associées au mot 1 (contexte 1), les points bleus correspondent aux phrases porteuses associées au mot 2 (contexte 2), les points rouges correspondent aux phrases porteuses sans lien avec aucun des deux mots (contexte 0).

Nous avons récemment appliqué cet outil¹⁶ dans le cadre d'un projet portant sur l'étude des mécanismes d'adaptation au contexte dans des situations dans lesquelles les variations contextuelles acoustiques et sémantiques peuvent interférer. Ce travail nous a permis de concevoir vingt-huit triplets de phrases associés à vingt-huit paires de mots phonétiquement proches et de quantifier les relations de proximité sémantique entre les mots de chaque phrase. Prenons, par exemple, la paire de mots « balle » / « belle ». Nous avons conçu trois portions de phrases sujet-verbe pour cette paire :

- ▶ Une portion qui est sémantiquement associée à « balle » (mais pas à « belle ») : « Le joueur a dévié la... » ;
- ▶ Une portion qui est sémantiquement associée à « belle » (mais pas à « balle ») : « Le prince a charmé la... » ;
- ▶ Une portion qui n'est associée à aucun des deux mots : « La salade a raccourci la... ».

Les résultats de la quantification obtenue à partir d'un modèle de langue issu des pages en français de Wikipédia sont représentés sur la figure 2. Les phrases correspondantes sont actuellement utilisées pour mettre en place un recueil de données sur l'adaptation perceptive à la variation.

Ces deux ensembles d'outils complètent avantageusement l'éventail des ressources computationnelles disponibles pour l'étude de la variation en perception de la parole par leur apport de solutions nouvelles. Nous les utilisons par exemple dans le cadre de deux projets collaboratifs en cours avec Étienne Gaudrain (Centre de Recherche en Neurosciences de Lyon, CRNL, UMR5292, CNRS / Inserm / Université Claude Bernard Lyon 1) et le Pr. Deniz Baskent (University Medical Center Groningen, UMCG, Pays-Bas). Ces deux projets reposent sur des manipulations des propriétés acoustiques et / ou sémantiques des stimuli linguistiques utilisés et ont pour objectif de nous aider à mieux comprendre, sous des angles différents, les mécanismes d'adaptation à la variation en perception de la parole. Ils ont reçu les soutiens respectifs de la Mission pour les Initiatives Transverses et l'Interdisciplinarité (MITI) du CNRS et du programme Marie Skłodowska-Curie (PRESTIGE-2017-2-0044, UE).

contact&info
 ▶ Olivier Crouzet
 LLING
olivier.crouzet@univ-nantes.fr

16. Leprieur L., Crouzet O. & Gaudrain E. 2020, Une base de données de phrases en français pour l'étude du rôle conjoint des incertitudes sémantique et acoustique dans la perception de la parole, in Benzitoun C., Braud C., Huber L., Langlois D., Ouni S., Pogodalla S. & Schneider S. (Éds.), 33èmes Journées d'Études sur la Parole (JEP), pp 353-361, Nancy, France : ATALA.

Modéliser pour comprendre : comment un programme de modélisation computationnelle des processus de perception et de production de la parole peut aider à spécifier la nature sensorielle et motrice des unités du langage

Directeur de recherche CNRS au sein du laboratoire *Grenoble Images Parole Signal Automatique (GIPSA-lab, UMR5216, CNRS / Université Grenoble Alpes)*, Jean-Luc Schwartz étudie la nature des processus cognitifs qui gouvernent la communication parlée. Il développe une approche intégrée des mécanismes de perception et de production de la parole, alliant études comportementales et neurocognitives chez l'adulte et l'enfant, analyse des systèmes linguistiques et modélisation computationnelle. Ces recherches interdisciplinaires ont débouché sur un cadre théorique novateur et des modèles capables de décrire les processus de développement et d'adaptation de la parole en situation d'interaction. Il a été, en 2017, lauréat de la médaille d'argent du CNRS.

Les sciences du langage butent depuis soixante ans sur une question importante, celle de la nature, sensorielle ou motrice, des unités du langage. Pour comprendre d'où vient cette question, il suffit de repartir de ce qu'est un acte de communication langagière. Parler permet à un locuteur de changer, à distance, l'état de connaissances, dans un sens très général, de son interlocuteur. Faute de pouvoir entrer directement dans la tête de l'auditeur, le locuteur émet des signaux acoustiques (du son) en bougeant ses « articulatoires » (langue, mâchoire, lèvres, voile du palais, larynx) et ces signaux sont captés par l'auditeur (Figure 1). De cerveau de locuteur à cerveau d'auditeur, le message linguistique est donc encodé et transmis par des « gestes articulatoires » et par des sons produits par ces gestes. Mais alors, est-ce que la nature du code phonologique sous-jacent — c'est-à-dire des unités dites « phonologiques », voyelles, consonnes, syllabes, formant les mots du langage oral — est définie par les propriétés des gestes produits par le locuteur ou des sons reçus par l'auditeur ?

Cette question a été posée dès les années 1950 dans la célèbre « Théorie Motrice » (TM) de la perception de la parole¹ qui postule que les unités de la phonologie (les phonèmes) seraient caractérisées par leurs propriétés articulatoires, et que l'auditeur serait capable de percevoir les actions motrices du locuteur à partir des sons perçus, pour décoder ces unités. Les théories auditives (TA) proposent au contraire que les unités doivent être définies en termes auditifs et non articulatoires ou moteurs.

Ces théories « sensorielles » et « motrices » se relient à la découverte par les neurosciences cognitives des « neurones miroir » chez le singe et des « systèmes miroir » chez les humains, aires corticales dédiées au départ au contrôle et à l'exécution des actions, et qui s'activent « en miroir » lorsque l'on perçoit un congénère effectuant ces mêmes actions. Ces systèmes d'inférence motrice, par lesquels le cerveau des primates — et particulièrement des humains — serait capable d'inférer les actions à partir de leurs conséquences sensorielles, fourniraient un support neuroanatomique plausible à la TM².

Alors, unités sensorielles ou unités motrices ? C'est là que peut entrer en scène la modélisation computationnelle. On peut, en effet, tenter de « modéliser pour comprendre ». Le modélisateur doit s'atteler à formuler de manière précise les hypothèses retenues ; à les formaliser en termes mathématiques rigoureux ; et, enfin, à tenter d'en dérouler les conséquences mathématiques sous forme de prédictions ou de tests d'hypothèses comparatives. La modélisation computationnelle dans cette acception apparaît comme un outil formel ou algorithmique au service d'une approche hypothético-déductive, dans laquelle on formule des hypothèses et on teste leur comportement, le plus souvent sous forme comparative entre jeux d'hypothèses concurrentes.

C'est cette voie que nous avons suivie depuis une dizaine d'années avec mes collègues Julien Diard, Pierre Bessière et Pascal Perrier, et nos jeunes collègues Clément Moulin-Frier, Raphaël Laurent, Marie-Lou Barnaud et Jean-François Patri. Nous avons entrepris de tester les propriétés comparatives des théories auditives et motrices de la parole, en « mettant en équation » le schéma de la Figure 1, sous forme d'un « modèle probabiliste bayésien », le modèle COSMO (pour « *Communicating Objects using Sensory-Motor Operations* »), présenté dans la Figure 2.



Figure 1 – Un acte de communication langagière véhiculé par des gestes articulatoires du locuteur et des sons captés par l'auditeur (d'après Barnaud L.J. 2018, *Modélisation bayésienne du développement conjoint de la perception, l'action et la phonologie*, Doctorat de l'Université de Grenoble.)

1. Liberman A.M. 1957, *Some results of research on speech perception*, *Journal of the Acoustical Society of America*, 29 : 117-123.

2. Schwartz J.L., Sato M. & Fadiga L. 2011, Le langage commun de la perception et de l'action dans la communication parlée : une perspective neurocognitive, *Faits de Langue*, 37 : 117-136.

Ce modèle repose sur une hypothèse fondamentale qui postule que le schéma de communication de la Figure 1 est « internalisé » dans le cerveau d'un agent communicant, à la fois locuteur et l'auditeur, et capable de simuler mentalement l'auditeur ou le locuteur avec lequel il communique. COSMO est un modèle probabiliste bayésien, qui décrit toutes les relations entre variables par des modèles probabilistes et déroule ses règles d'inférence pour déterminer comment passer de la connaissance sur une ou plusieurs variables à la connaissance sur d'autres variables du modèle. Deux variables décrivent l'objet de communication (concept, mot, syllabe ou phonème), selon qu'il active la branche motrice, celle du « *Speaker* » O_s , ou la branche sensorielle, celle du « *Listener* » O_l . À ces variables linguistiques sont associées des variables biophysiques, qui décrivent des actions et gestes articulatoires (M) ou des caractéristiques auditives, voire visuelles ou tactiles (S). Une variable C connecte O_s et O_l , prenant la valeur 1 si $O_s = O_l$ et 0 dans le cas contraire.

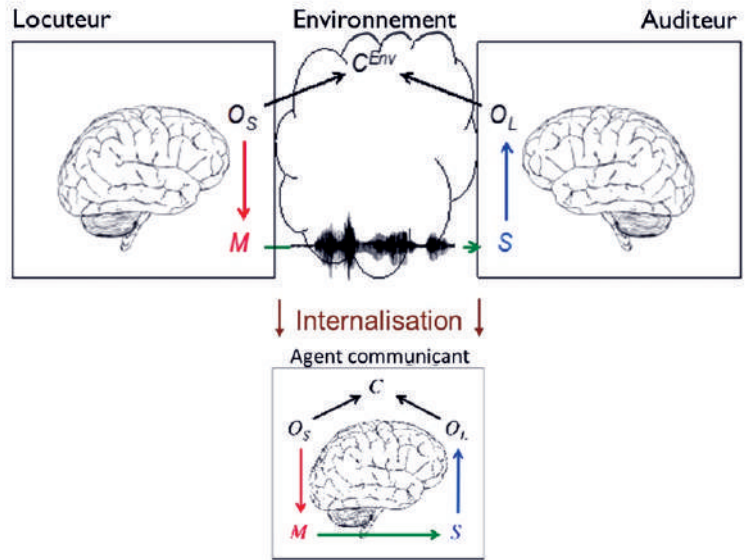


Figure 2 – Le modèle COSMO « *Communicating Objects using Sensory-Motor Operations* » (d'après Laurent R., Barnaud M.-L., Schwartz J.-L., Bessièrè P. & Diard J. 2017, *The complementary roles of auditory and motor information evaluated in a Bayesian perceptuo-motor model of speech perception*, *Psychological Review*, 124 : 572-602 et Barnaud L.J. 2018, *Modélisation bayésienne du développement conjoint de la perception, l'action et la phonologie*, Doctorat de l'Université de Grenoble)

L'agent est alors entièrement défini par la loi de probabilité reliant ces cinq variables — qui incidemment fournissent également l'acronyme du modèle COSMO. Cette loi de probabilité est décomposée et simplifiée selon la Figure 3, dans laquelle apparaissent des termes correspondant à des modèles cognitifs classiques : probabilité de chaque objet (« *prior* »), répertoire moteur $P(M / O_s)$ associant les gestes moteurs adéquats à un objet donné pour le « *Speaker* » ; modèle interne $P(S / M)$ associant dans le cerveau de l'agent le son correspondant à un geste donné ; classifieur auditif $P(O_l / S)$ associant l'objet pour le « *Listener* » à un son donné ; et système de validation $P(C / O_s O_l)$.

consiste à inférer les gestes articulatoires produits par le locuteur (*Speaker*). L'inférence bayésienne $P(O_s | S) \propto \sum_M P(M | O_s)P(S | M)$ fournit alors la réponse, et donc la simulation computationnelle quantitative de la TM, passant par un processus dit « d'inversion articulatoire-acoustique » permettant d'inférer le geste M correspondant au son S , puis par un processus de décodage articulatoire permettant de trouver l'objet O_s , par exemple le phonème, correspondant au geste inféré M .

Ce modèle a l'intérêt considérable de permettre de tester et de comparer théories auditives et motrices de la perception de la parole³. En effet, dans COSMO, une tâche perceptive consiste à chercher à inférer l'objet à partir du son, donc à répondre à la « question probabiliste » $P(O / S)$. Mais l'existence de deux objets, O_s et O_l , fournit un lien immédiat vers les deux jeux de théories. Pour les TA, c'est O_l qui fournit le pivot du décodage perceptif, et la réponse correspond directement au classifieur auditif $P(O_l / S)$. Pour la TM, c'est O_s qui fournit le pivot, puisque percevoir

Ainsi, COSMO fournit, pour la toute première fois, un cadre intégrateur permettant de comparer les prédictions des théories auditives et motrices de la perception (et de la production) de la parole dans un ensemble de situations expérimentales. Or, nos premiers travaux ont conduit à un premier résultat très surprenant et, à notre sens, extrêmement important. Dans certaines conditions « idéales » d'apprentissage des modèles cognitifs élémentaires $P(M / O_s)$, $P(O_l / S)$ et $P(S / M)$, on peut montrer que les prédictions des théories auditives $P(O_l / S)$ et motrices $P(O_s / S)$

$$P(C O_s S M O_l) = P(O_s) \times P(M | O_s) \times P(S | M) \times P(O_l | S) \times P(C | O_s O_l)$$

Le diagramme illustre la décomposition de l'équation probabiliste. Les termes de l'équation sont alignés au-dessus de cinq boîtes à l'acronyme COSMO. Des flèches colorées relient chaque terme à sa boîte correspondante : une flèche noire pour $P(O_s)$ (Prior sur les objets), une flèche rouge pour $P(M | O_s)$ (Répertoire moteur), une flèche verte pour $P(S | M)$ (Modèle interne), une flèche bleue pour $P(O_l | S)$ (Classifieur auditif), et une flèche noire pour $P(C | O_s O_l)$ (Système de validation).

Figure 3 – L'équation probabiliste du modèle COSMO et sa décomposition (d'après Laurent R., Barnaud M.-L., Schwartz J.-L., Bessièrè P. & Diard J. 2017, *The complementary roles of auditory and motor information evaluated in a Bayesian perceptuo-motor model of speech perception*, *Psychological Review*, 124 : 572-602 et Barnaud L.J. 2018, *Modélisation bayésienne du développement conjoint de la perception, l'action et la phonologie*, Doctorat de l'Université de Grenoble)

3. Barnaud M.L., Bessièrè P., Diard J. & Schwartz J.L. 2018, *Reanalyzing neurocognitive data on the role of the motor system in speech perception within COSMO, a Bayesian perceptuo-motor model of speech communication*, *Brain & Language*, 187: 19-32.

sont en réalité parfaitement identiques⁴ ! Ce résultat, que nous avons baptisé « théorème d'indistingabilité », a la beauté — et les limites — d'un résultat mathématique : il est parfaitement exact mais limité à des conditions idéales d'apprentissage qui ne sont en réalité jamais remplies. Mais il illustre parfaitement pourquoi les débats entre TA et TM sont si complexes depuis soixante ans : en réalité, c'est dans l'écart avec les conditions parfaites du théorème d'indistingabilité qu'il faut chercher les possibilités de contraster les deux ensembles de théories.

On peut alors tenter de comparer les propriétés respectives de la « branche auditive » de décodage $P(O_L / S)$ — correspondant à l'implémentation de la TA — et de la « branche motrice » de décodage $P(O_S / S)$ — correspondant à l'implémentation de la TM. Nous avons pu montrer que, en situation réelle, la « branche auditive » fournit un décodage plus efficace de sons perçus lorsque ces sons correspondent aux conditions d'apprentissage — lorsqu'ils sont typiques et non bruités. La « branche motrice » est, elle, plus robuste en situation de stimuli atypiques, bruités ou correspondant par exemple à des sons émis par des locuteurs étrangers. Ces résultats sont conformes aux données de neuroimagerie, avec une activité renforcée des aires corticales motrices — correspondant à la mise en œuvre de la branche motrice de perception — pour des stimuli bruités ou atypiques.

Dans COSMO, on peut proposer que la perception ne soit en réalité ni purement auditive ni purement motrice, mais qu'elle repose sur une combinaison efficace, une « fusion » des voies auditive et motrice, plus efficace dans toutes les situations de fonctionnement. Ceci correspond à une mise en équation d'une théorie sensori-motrice de la communication parlée que nous avons proposée dans les années 2000⁵. COSMO permet alors de faire des prédictions multiples en bon accord avec les données expérimentales :

- ▶ sur la possibilité de modifier un percept auditif de parole par une action somatosensorielle sur le visage, activant la voie motrice de la perception ;
- ▶ sur la possibilité pour un enfant de discriminer dès la naissance tous les sons des langues, grâce à la voie auditive mature précocement, mais d'accéder à l'invariance des consonnes seulement lorsqu'il est capable de commencer à babiller et ainsi d'activer sa voie motrice de perception.

Enfin, COSMO fournit des prédictions pertinentes sur la structure des systèmes sonores des langues du monde, au confluent de contraintes sensorielles et motrices, et permettant de comprendre pourquoi certaines voyelles et certaines consonnes sont présentes dans presque toutes les langues du monde⁶.

On le voit, la modélisation computationnelle apparaît ici comme un formidable outil de formalisation, de comparaison et de test d'hypothèses, capable de produire des prédictions expérimentales, et d'éclairer notre compréhension de la structure de la parole et de la nature des unités du langage, ouvrant la voie à un programme d'expérimentation de formalisation linguistique tout à fait prometteur.

contact&info

▶ Jean-Luc Schwartz
GIPSA-lab

Jean-Luc.Schwartz@gipsa-lab.
grenoble-inp.fr

4. Laurent R., Barnaud M.-L., Schwartz J.-L., Bessière P. & Diard J. 2017, *The complementary roles of auditory and motor information evaluated in a Bayesian perceptuo-motor model of speech perception*, *Psychological Review*, 124 : 572-602.

5. Schwartz J.L., Basirat A., Ménard L. & Sato M. 2012, *The Perception for Action Control Theory (PACT): a perceptuo-motor theory of speech perception*, *Journal of NeuroLinguistics*, 25 : 336-354.

6. Moulin-Frier C., Diard J., Schwartz J.L. & Bessière P. 2015, *COSMO ("Communicating about Objects using Sensory-Motor Operations"): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems*, *Journal of Phonetics*, 53 : 5-41.



Littérature et Révolution

Le carnet *Littérature et Révolution* est une publication d'Olivier Ritz, maître de conférences en littérature française à l'université de Paris. Depuis 2016, il y interroge le traitement de la Révolution française par la littérature et le théâtre. Comment cet événement historique est-il mis en scène et représenté ? Les [comptes rendus critiques](#) d'œuvres constituent le moyen privilégié par l'auteur pour partager, sur son carnet, une partie de son travail de recherche. Les lecteurs et lectrices sont ainsi invités à découvrir l'utilisation — et ses limites — de la Révolution comme toile de fond historique dans des romans tels que *Les Talons rouges* ou encore *La Sonate à Bridgetower*. Ils pourront également lire les commentaires de l'auteur sur deux ouvrages de Jean-Clément Martin visant à dresser le bilan des [légendes](#) et des [approximations](#) autour de la période qualifiée de « Terreur ». Le théâtre est aussi bien représenté dans les comptes rendus proposés. Plusieurs spectacles sont longuement commentés dont l'opéra-comique *Richard Cœur de Lion* ainsi que, par exemple, les pièces *J'ai rêvé la Révolution* et *La Mort de Danton*.

Le corpus abordé tout au long de *Littérature et Révolution* se distingue donc par son éclectisme et par l'important travail de sélection et d'organisation des matériaux. À cette sélection, s'ajoutent des analyses de textes variés — comme l'illustre ce billet sur le poème « Mauprat » de George Sand — et des [ressources](#) précieuses pour la recherche. Alors que la Révolution française semble être un sujet prisé, y compris dans la culture populaire comme peut par exemple le faire penser la création d'une série télévisée à son sujet, ce carnet invite de façon bienvenue à une lecture attentive et critique de la représentation qui est faite de cet événement historique, en mobilisant habilement tous les outils de l'analyse littéraire.

Céline Guilleux

Littérature et Révolution Comptes rendus Recherches Ressources Annonces Qui suis-je ? Crédits
Carnet de recherche d'Olivier Ritz

Sand, Mauprat et le Roi des Aulnes

11 novembre 2020
Recherches, Réflexions
Allemagne, Fiction, Sand
Laisser un commentaire
Modifier

« Je me trouvais donc orphelin à sept ans. Mon grand-père pillait dans la maison de ma mère tout l'argent et les nippes qu'il put emporter ; puis, laissant le reste, et disant qu'il ne voulait point avoir affaire aux gens de loi, il n'attendit pas que la morte fût ensevelie, et, me prenant par le collet de ma veste, il me jeta sur la croupe de son cheval. »
— Édition Folio, p. 43-44.

RECHERCHE... [🔍]

COMMENTAIRES RÉCENTS

- 16 de juillet 1794 : un acte héroïque dans *Quand la littérature donne des leçons au peuple*
- La tragédie *Tommy de Marie-Anne Barbier* en ligne numérique | Cité des Dames. Créatives dans la cité dans *Un atelier d'édition collaborative*
- Filmer la Révolution aujourd'hui : rencontre avec Pierre Schoeller — IMARIV 18-21 dans *Pierre Schoeller à Paris Diderot*
- Entres et Champs-Élysées — IMARIV 18-21 dans *Entres et Champs-Élysées*
- La Femme grondalet, 2e édition ! — *Littérature et Révolution dans un atelier d'édition collaborative*

ARTICLES RÉCENTS

- Sand, Mauprat et le Roi des Aulnes
- La Révolution cataloguée
- Des écrits qui agissent
- La Terreur : naissance d'un théâtre d'ombres
- Lettres de la Vendée : un roman du consentement

MOTS-CLEFS

14 juillet | 19e siècle | Actusite
Allemagne | Anthologie | Bastille

contact&info

▶ Olivier Ritz
oritz@orange.fr

▶ Pour en savoir plus

<https://litrev.hypotheses.org>

<https://www.openedition.org/16797>

contact&info

▶ Céline Guilleux

celine.guilleux@openedition.org

OpenEdition

▶ Pour en savoir plus

<https://www.openedition.org>

la lettre de l'InSHS

- ▶ **Directeur de la publication** François-Joseph Ruggiu
- ▶ **Directrice de la rédaction** Marie Gaille
- ▶ **Responsable éditoriale** Armelle Leclerc armelle.leclerc@cnrs-dir.fr
- ▶ **Conception graphique** Sandrine Clérisse & Bruno Roulet, Secteur de l'imprimé PMA
- ▶ **Graphisme Bandeau** Valérie Pierre, direction de la Communication CNRS
- ▶ **Crédits images Bandeau**
© Photothèque du CNRS / Hervé Théry, Émilie Maj, Caroline Rose, Kaksonen
- ▶ **Pour consulter la lettre en ligne**
www.cnrs.fr/inshs/Lettres-information-INSHS/lettres-informationINSHS.htm
- ▶ **S'abonner / se désabonner**
- ▶ **Pour accéder aux autres actualités de l'InSHS**
www.cnrs.fr/inshs
- ▶ **Retrouvez l'InSHS sur Twitter** @INSHS_CNRS

Institut des sciences humaines et sociales CNRS

• 3 rue Michel-Ange 75794 Paris cedex 16 •

ISSN : 2272-0243