

Avant-propos

À l'origine de ce livre : le projet ès lettres

Cécile Obligi
Laurie Aoustet

Ès lettres est un projet d'étude et de valorisation des thèses de doctorat ès lettres soutenues en France au XIX^e siècle. Il part du constat, déjà fait par Émile Durkheim, que c'est la délivrance des grades, c'est-à-dire la mise en place d'un contrôle juridiquement certifié d'un corpus donné de savoirs et de compétences, qui distingue fondamentalement l'université des institutions d'enseignement qui l'ont précédée. Or, au sein du système universitaire tel qu'il se cristallise dans le courant du XIX^e siècle, l'un de ces grades, le doctorat, acquiert une importance spécifique, d'une part comme certifiant la capacité à produire des savoirs nouveaux, d'autre part comme épreuve régulant l'accès au corps universitaire lui-même. Le doctorat se trouve ainsi à l'interface entre le système de production des savoirs scientifiques et le système de reproduction des élites intellectuelles, et constitue un observatoire de l'institutionnalisation des disciplines. Et pourtant, il reste à ce jour très peu exploré par l'historiographie, par manque d'outils pour aborder une documentation à la fois massive et dispersée – et aussi parce que la mémoire universitaire a eu tendance à dévaloriser les doctorats les plus anciens, sans nécessairement prendre la peine de les lire. Partant de ce constat, le projet ès lettres vise tout d'abord à rassembler des informations relatives aux thèses et actuellement dispersées entre de multiples sources, puis à procéder à la numérisation de ces thèses tout en préparant celle de documents qui leur sont associés (notamment les rapports de soutenance). Il s'agit en outre d'élaborer une bibliographie à la fois générale et spécialisée sur ces thèses, de constituer une base de données en ligne et reliée à des référentiels à partir de l'ensemble de ces éléments, et enfin de valoriser scientifiquement ces matériaux.

Principes généraux

Le projet s'appuie en premier lieu sur la collection des thèses de doctorat ès lettres conservée par la Bibliothèque interuniversitaire de la Sorbonne (BIS), quasi exhaustive jusqu'en 1939 au moins¹ : non seulement la faculté des lettres de Paris décerne quatre grades sur cinq au XIX^e siècle, mais sa bibliothèque a hérité de la collection dite « ministérielle ». Celle-ci s'est constituée à la suite d'une circulaire du 8 juin 1811 qui oblige les candidats² à envoyer deux exemplaires de leur thèse au Conseil de l'Université, puis à son héritier le ministère de l'Instruction publique ; or ces institutions, n'ayant pas pour mission de conserver durablement de massives collections d'ouvrages académiques, ont très tôt choisi d'en faire don à la bibliothèque de la Sorbonne. En 1841, le Conseil royal de l'Instruction publique, soucieux d'enrichir les bibliothèques attachées aux facultés non parisiennes, décide qu'en sus des exemplaires de thèse déjà exigés, tout candidat devra en déposer un nombre égal à celui des autres facultés des lettres du royaume³. Cependant, le centralisme reste de règle, et toutes ces pièces doivent d'abord être adressées au ministre, ce dernier se chargeant de la redistribution⁴. À l'heure actuelle, le corpus comprend 2 125 thèses pour la période 1810-1899⁵ (dont 1 704 soutenues à Paris), ce qui correspond à la première phase du projet (2020-2023), financée notamment grâce à l'appel à projets CollEx-Persée 2019-2020⁶.

La valorisation de ce corpus passe par la collaboration entre bibliothécaires, archivistes, chercheuses et chercheurs au niveau national, au sein d'un

1 On trouvera une description de ce fonds dans Jean-Pierre CALVIGNAC, « Les thèses : un fonds majeur de la bibliothèque de la Sorbonne », dans Claude JOLLY et Bruno NEVEU (dir.), *Éléments pour une histoire de la thèse*, op. cit., p. 177-249.

2 Y compris ceux qui finalement échouent : le corpus comprend donc des thèses – une trentaine – qui ont été refusées par le jury. Voir la « Circulaire portant invitation aux recteurs d'adresser deux exemplaires de toutes les thèses soutenues devant les facultés », *Recueil des lois et règlements concernant l'Instruction publique depuis l'édit de Henri IV, en 1598, jusqu'à ce jour*, t. 5, Paris, Brunot-Labbe, 1820, p. 276.

3 « Arrêté qui fixe le nombre et la répartition des exemplaires des thèses de doctorat, 7 décembre 1841 », Arthur de BEAUCHAMP, *Recueil des lois et règlements sur l'enseignement supérieur*, t. 1, Paris, Delalain frères, 1880, p. 917-918.

4 On trouvera plus d'informations sur la distribution des thèses dans les bibliothèques universitaires sur le site de l'exposition virtuelle du projet, <https://nubis.univ-paris1.fr/s/theses-doctorats-es-lettres-19-siecle-exposition-devenir-savant/page/page-2> [consulté le 28/05/2024].

5 Le *terminus ad quem* de 1899 s'explique par la soutenance, cette année-là, des premières thèses de doctorat d'université, nouveau titre créé en 1897 dans le but d'attirer les étudiants étrangers : d'autres questions s'ajoutent alors, que nous avons réservées pour une phase ultérieure du projet.

6 *Ès lettres. Corpus des thèses de doctorat ès lettres en France au XIX^e siècle*, CollEx-Persée, <https://www.collexperse.eu/projet/es-lettres> [consulté le 28/05/2024].

projet porté par une conservatrice des bibliothèques, Cécile Obligi, et un enseignant-chercheur, Pierre Verschueren. À la BIS, les principaux aspects du projet sont supervisés par le service de la valorisation numérique des collections et du soutien à la recherche (SERVAL), dirigé par Cécile Obligi, assistée de son adjointe Laurie Aoustet, unité spécialement créée pour soutenir les projets en humanités numériques. Les chercheurs et chercheuses sont impliqués par le biais d'un comité scientifique pluridisciplinaire de vingt et un membres, historiens et historiennes, mais aussi spécialistes de sociologie, de philosophie ou de littérature, et un comité de valorisation de neuf membres, coordonné par Arnaud Desvignes. En outre, quatre laboratoires de recherche sont partenaires : le Centre Lucien Febvre (université Marie et Louis Pasteur), l'Institut d'histoire moderne et contemporaine (UMR 8066, sous la tutelle du Centre national de la recherche scientifique, de l'École normale supérieure et de l'université Paris 1 Panthéon-Sorbonne), le Centre d'histoire du XIX^e siècle (université Paris 1 Panthéon-Sorbonne et Sorbonne Université) et le Centre d'histoire culturelle des sociétés contemporaines (université de Versailles Saint-Quentin-en-Yvelines). Du côté des établissements documentaires, le projet fait appel à une collaboration avec la Bibliothèque nationale de France, les Archives nationales, mais aussi avec la bibliothèque de l'École nationale des chartes, les services communs de documentation de l'université Marie et Louis Pasteur, d'Aix-Marseille Université et de l'université de Toulouse Jean-Jaurès – liste qui est appelée à s'étendre, à mesure que les universités voudront se saisir de l'outil pour valoriser leurs collections.

Stratégie

Le premier volet du projet entend rassembler un grand nombre d'informations relatives aux thèses de doctorat et à leurs auteurs, informations actuellement dispersées entre de multiples sources : Archives nationales (en particulier les séries AJ/16 et F/17), archives départementales, BIS, voire fonds privés. Cette tâche a été confiée à Lucie Lachenal, ingénieure d'études salariée grâce au financement du GIS CollEx-Persée, puis à Ève Givois et Malika Khobeizi. Concernant les thèses elles-mêmes, sont relevées les données suivantes : titre, langue, éventuelle dédicace, lieu d'édition, date d'édition, lieu d'impression, date d'impression, date de validation par le doyen, date d'approbation (autorisation d'impression donnée par le recteur), pagination, format, cote du document à la BIS, auteur commenté (en cas d'édition scientifique), nom du doyen ayant validé, nom du recteur ayant approuvé, rapporteur éventuel⁷, nom

7 L'article 2 de l'arrêté du conseil royal du 17 juillet 1840 prévoit que « les thèses manuscrites seront remises au doyen, qui les fera examiner par le professeur chargé de l'enseignement auquel chaque

des membres du jury, dédicataire, personnes remerciées, éditeur ou maison d'édition (nom, adresse, ville), imprimeur ou imprimerie (nom, adresse, ville), établissement de soutenance, mention, obtention de l'unanimité ou non. Concernant les auteurs : nom, prénoms, date de naissance, lieu de naissance, date et lieu d'obtention de la licence, titres universitaires tels que l'agrégation ou le titre d'ancien élève de grande école, éléments de carrière. Sur ce dernier point, le projet reprend et étend les travaux lancés à l'initiative de Françoise Huguet au service d'histoire de l'éducation de l'Institut national de la recherche pédagogique (INRP)⁸.

Le second volet correspond à une opération de numérisation et océrisation de l'ensemble du corpus des thèses, pour valoriser le fonds et encourager son étude. Cette opération est, pour plus de la moitié, assurée par la Bibliothèque nationale de France, par le biais de la « filière partenaires » de son marché de numérisation. Les thèses sont progressivement mises en ligne sur Gallica, ce qui garantit leur conservation pérenne et améliore leur visibilité, et sont regroupées à l'intérieur d'une collection spécifique, de manière à faciliter les recherches et la fouille de texte⁹. Les thèses du XIX^e siècle sont destinées à être toutes numérisées, mais étant donné l'ampleur du chantier, n'ont été traitées de la sorte, dans un premier temps, qu'une partie des thèses parisiennes, pour un total d'environ 200 000 pages. Le reste des thèses sera numérisé par des prestataires, dans le cadre d'un marché de numérisation mis en place par la BIS. Il est en outre prévu de numériser une série de documents associés, en particulier les rapports de soutenance qui peuvent être retrouvés aux Archives nationales¹⁰, mais aussi une partie au moins des registres de procès-verbaux de soutenance, certains jurys ne pouvant être reconstitués qu'à partir des paraphes apposés au bas de ces documents.

thèse se rapporte. Celui-ci donnera son avis sur l'admissibilité de la thèse » (« Arrêté portant règlement relatif aux examens du doctorat ès lettres, 17 juillet 1840 », Arthur de BEAUCHAMP, *Recueil des lois et règlements sur l'enseignement supérieur*, op. cit., p. 869). C'est l'acte de naissance de la pratique du rapport de thèse. L'usage semble s'être progressivement affirmé de prendre contact avec le futur rapporteur très tôt, dans l'économie d'une thèse, ne serait-ce que pour s'assurer de la recevabilité du sujet ; des relations de patronage se construisent alors, que l'arrêté de 1969 inscrit dans le droit – tout en exigeant deux rapports supplémentaires pour autoriser la soutenance.

- 8 Françoise HUGUET, « Les thèses de doctorat ès lettres soutenues en France de la fin du XVIII^e siècle à 1940 », Ressources numériques en histoire de l'éducation, novembre 2009, <http://rhe.ish-lyon.cnrs.fr/?q=theses> [consulté le 28/05/2024].
- 9 La collection des thèses de doctorat ès lettres conservées par la Bibliothèque interuniversitaire de la Sorbonne (BIS) et mises en ligne sur Gallica est accessible par son « code projet », « BisThLet », à taper entre guillemets dans le moteur de recherche de Gallica.
- 10 Au moment de publier cet article, sont déjà accessibles les rapports de la période 1890-1901, conservés aux Archives nationales (AN), F/17/13249, https://www.siv.archives-nationales.culture.gouv.fr/siv/UD/FRA_IR_041030/d_2_285_19_2 [consulté le 28/05/2024].

À partir de ces éléments, une base de données relationnelle a été constituée par Alyx Taounza-Jeminet, ingénieure en humanités numériques au Serval jusqu'en août 2022, puis mise en ligne grâce à la création d'un site web par Sébastien Clément, ingénieur de recherche à la BIS. Une interface de recherche permet aux chercheurs et chercheuses comme aux simples curieux et curieuses de retrouver facilement les informations recueillies et leur source¹¹. Elle donne accès, par des liens, aux images numérisées des thèses et documents associés (portraits de thésards et de professeurs, notices biographiques, articles de périodiques...) stockées sur diverses plateformes. La base de données elle-même, construite en langage de requête structuré (SQL), a été élaborée grâce à l'outil Heurist¹², service gratuit en *open source* recommandé par l'infrastructure de recherche Huma-Num. Ce logiciel permet, au moyen d'un navigateur web et sans programmation, de structurer, gérer, analyser, visualiser, publier et archiver des données riches, hétérogènes et interconnectées ; la structure de la base de données peut être modifiée et étendue au fur et à mesure de l'évolution des besoins. La durabilité des bases de données Heurist est assurée par une structure MySQL sous-jacente documentée et stable, une communauté d'utilisateurs et utilisatrices active, un hébergement partagé sur un petit nombre de serveurs et un point de maintenance unique.

La base ès lettres est reliée à des référentiels, en particulier Wikidata et la plateforme IdRef, gérée par l'Agence bibliographique de l'enseignement supérieur (Abes), qui fournit des identifiants pérennes, construits sur l'identifiant de la notice Sudoc, ce qui permet d'identifier sans ambiguïté les personnes (physiques et morales), les lieux, les concepts et les œuvres liés à une ressource donnée. Par ce moyen, les données consolidées pourront être à terme reversées en masse dans les notices bibliographiques du Sudoc¹³. La base de données est aussi reliée à d'autres projets et ressources, comme le Dictionnaire

11 Ès lettres. Les thèses de doctorat ès lettres en France au XIX^e siècle, <https://eslettres.bis-sorbonne.fr> [consulté le 28/05/2024].

12 Expert Nation est un bon exemple d'une base de données sous Heurist présentant les parcours et circulations d'universitaires, en l'occurrence des Australien-nes du début du XX^e siècle, <https://heuristnetwork.org/expert-nation> [consulté le 24/07/2024].

13 Le catalogue collectif Sudoc est une base de données bibliographiques nationale et l'on y trouve de fait la quasi-totalité des notices de thèses ès lettres soutenues en France au XIX^e siècle. Mais en raison de l'absence d'un certain nombre d'informations importantes sur les thèses de cette époque, de la multiplicité des contributeurs et des délais contraints dans lesquels se fait parfois le travail de catalogage, les données sont assez lacunaires, voire erronées.

des éditeurs français du XIX^e siècle (DEF19)¹⁴, la France savante¹⁵, annuaire prosopographique des membres de sociétés savantes créé par le Comité des travaux historiques et scientifiques (CTHS), les éventuels fonds signalés dans le catalogue Calames (en particulier archives privées et correspondances), les dictionnaires biographiques publiés par l'INRP et mis en ligne sur la Bibliothèque historique de l'éducation¹⁶. Elle indiquera à terme les éventuels articles de journaux et de revues consacrés aux soutenances disponibles sur Retronews, Persée et Gallica. Pour permettre la meilleure intégration possible de ces données liées issues de diverses ressources, ainsi que leur interopérabilité, le modèle de données est conçu selon les normes FRBR (pour *Functional Requirements for Bibliographic Records*, soit « fonctionnalités requises des notices bibliographiques ») de représentation des documents¹⁷. La base permet d'exporter les données dans plusieurs formats. L'ensemble peut ainsi être réutilisé pour créer des bases comparables dans des régions particulières, dans d'autres disciplines ou à des périodes ultérieures.

Enfin, le projet ès lettres entend encourager la recherche sur les thèses par la mise en place d'un ambitieux programme de valorisation scientifique. Dans le cadre du projet, une journée d'étude a été organisée à Besançon le 23 mai 2022, suivie d'un colloque international qui a eu lieu à la Sorbonne les 8 et 9 septembre 2022, rassemblant à la suite d'un large appel à contributions les meilleur-es spécialistes de la question¹⁸. Cet ouvrage rassemble la plupart des communications produites lors de ces deux manifestations. En parallèle de ces manifestations, le corpus est valorisé par la rédaction de notes sur un carnet créé sur la plateforme Hypothèses¹⁹ et par la réalisation d'une exposition virtuelle, accessible sur la bibliothèque numérique de la BIS, NuBIS, et intitulée *Devenir savants : thèses et doctorats ès lettres au XIX^e siècle*²⁰. Les

14 DEF 19. Dictionnaire des éditeurs français du XIX^e siècle, <https://heurist.huma-num.fr/DEF19/web/37157> [consulté le 24/07/2024].

15 Annuaire prosopographique, la France savante, <https://cths.fr/an/prosopographie.php> [consulté le 23/07/2024].

16 Bibliothèque historique de l'éducation, <https://education.persee.fr> [consulté le 28/05/2024].

17 L'enjeu de la représentation FRBR se retrouve par exemple dans la relation entre le Sudoc et theses.fr. Voir le billet « theses.fr et les FRBR », *Punktokomo. Le blog technique de l'Agence bibliographique de l'enseignement supérieur*, 7 juin 2013, <https://punktokomo.abes.fr/2013/06/07/theses-fr-et-les-frbr> [consulté le 28/05/2024].

18 On trouvera le texte de l'appel et la composition du comité scientifique sur le carnet Hypothèses du projet ès lettres, <https://eslettres.hypotheses.org/278> [consulté le 28/05/2024].

19 Carnet Hypothèses du projet ès lettres, <https://eslettres.hypotheses.org> [consulté le 28/05/2024].

20 *Devenir savants : thèses et doctorats ès lettres au XIX^e siècle*, exposition virtuelle, <https://nubis.univ-paris1.fr/s/theses-doctorats-es-lettres-19-siecle-exposition-devenir-savant/page/introduction> [consulté le 28/05/2024].

données collectées dans le cadre du projet, en permettant de rendre compte de l'histoire des thèses de doctorat et de leur contexte de production, peuvent de fait nourrir bien des réflexions et devraient attirer l'attention de chercheurs et chercheuses travaillant sur des thèmes allant de l'histoire des sciences humaines et sociales (sociabilités savantes, émergence progressive de la sociologie, de la psychologie, de l'archéologie, etc.) à celle de l'enseignement supérieur (à la fois comme système de carrières et comme dispositif de production, de sélection et de hiérarchisation des savoirs et des savants), en passant par l'histoire de l'édition scientifique, celle des élites intellectuelles locales, ou encore celle de la réception des travaux universitaires à l'étranger (les premières recherches bibliographiques montrent que les thèses françaises ont donné lieu à des recensions régulières dans les revues anglophones et germanophones).

La création d'une collection spécifique regroupant les thèses numérisées ouvrira la voie à des recherches en texte intégral limitées à ce seul corpus et, au-delà, à des expériences de *text mining* – c'est-à-dire de recherche intelligente au sein du corpus océrisé afin, par exemple, d'évaluer la récurrence de mots, noms, sujets, ou d'analyser la structuration des textes. Des financements sont recherchés pour poursuivre le chantier de manière à traiter la période allant de 1899 à 1939, année marquant la fin de l'exhaustivité des collections de la BIS, les thèses non parisiennes n'étant, semble-t-il et sous réserve d'inventaire, plus systématiquement reçues au-delà de cette date.